

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 8, 2003, 11:03:24 ; Search time 41 Seconds
(without alignments)
963.974 Million cell updates/sec

Title: US-09-987-755-2

Perfect score: 1319

Sequence: 1 MAAVSLRLGLDVLWGKLRYP.....VCYQAITKKIKICEDLLLP 249

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A Geneseq_19Jun03.*

1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1319	100.0	249	AAU09404	Human hepatoma-der
2	1319	100.0	260	AAU20553	Human secreted pro
3	1319	100.0	260	AAU21747	Novel human neopla
4	1314	99.6	250	AAU21590	Novel human neopla
5	1291	97.9	547	AAW69240	Clone A073_3 prote
6	1291	97.9	547	AAW90720	Human A073_3 proc
7	1291	97.9	553	AAO16415	Human nucleic acid
8	1278	96.9	550	AAU21690	Novel human neopla
9	1263	95.8	269	AAU20515	Human secreted pro

10	1263	95.8	269	22	AAU21813	Novel human neopla
11	193	14.6	475	22	ABB58603	Drosophila melanog
12	190.5	14.4	202	23	ABB72327	Rat protein isolat
13	189	14.3	203	20	AAU28502	HGFH2 Human Growth
14	189	14.3	203	21	AAU13521	Human hepatoma-der
15	189	14.3	203	21	AAU88374	Human type II hepa
16	189	14.3	203	22	AAU92800	Human protein sequ
17	189	14.3	203	22	AAU70073	Human secreted pro
18	189	14.3	203	23	ABG65515	Human albumin fusi
19	186.5	14.1	246	21	AAU58173	Lung cancer associ
20	185.5	14.1	237	18	AAU33216	Mouse hepatoma der
21	185.5	14.1	240	16	AAU66727	Human hepatoma der
22	185.5	14.1	240	21	AAU09984	Human HDGFP prote
23	185.5	14.1	240	22	ABB50263	Hepatoma-derived g
24	178.5	13.5	235	19	AAU61963	Lung growth factor
25	166.5	12.6	326	22	ABB12116	Human PRO1604 homo
26	165	12.5	314	22	AAU70075	Human secreted pro
27	165	12.5	314	23	ABG65517	Human albumin fusi
28	165	12.5	671	21	AAU99426	Human PRO1604 (UNQ
29	165	12.5	671	22	AAU29196	Human PRO polypept
30	165	12.5	671	22	AAU66175	Protein of the inv
31	165	12.5	671	23	AAU97038	Human LP protein L
32	165	12.5	671	23	ABU05042	Human NOV4d protei
33	165	12.5	671	24	ABU71284	Human PRO1604 prot
34	165	12.5	671	24	ABU65741	Human secreted/tra
35	165	12.5	671	24	ABU66074	Novel human secret
36	165	12.5	671	24	ABU67578	Human secreted/tra
37	165	12.5	671	24	ABU65436	Human PRO polypept
38	165	12.5	671	24	ABU58572	Human PRO polypept
39	165	12.5	671	24	ABU56108	Human secreted/tra
40	165	12.5	671	24	ABU57103	Human PRO polypept
41	165	12.5	671	24	ABU10682	Human secreted/tra
42	165	12.5	676	22	AAU47218	Human NOV9 protein
43	165	12.5	676	22	AAU64373	Human hepatome cel
44	164	12.4	667	23	ABU05040	Human NOV4b protei
45	164	12.4	667	23	ABU05041	Human NOV4c protei

ALIGNMENTS

RESULT 1

AAU09404

ID AAU09404 standard; Protein; 249 AA.

XX AC AAU09404;

XX DT 20-MAY-1997 (first entry)

XX DE Human hepatoma-derived growth factor (HDGF-2).

XX DE Hepatoma-derived growth factor-2; HDGF-2; chemokine; wound healing;

XX DE vulnerable; burn; ulcer; thrombosis; arteriosclerosis; antibody;

XX DE antagonist; tumour; therapy.

XX OS Homo sapiens.

XX PN WO9639485-A1.

XX PD 12-DEC-1996.

XX PF 05-JUN-1995; 95WO-US06731.

XX PR 05-JUN-1995; 95WO-US06731.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Kunsch CA, Rosen CA;

XX DR WPI; 1997-043108/04.

XX DR N-PSDB; AAT47520.

XX PT DNA encoding human-derived growth factor polypeptide - useful to

```

PT promote wound healing as result of burns or ulcers
XX Claim 7; Fig 2; 54pp; English.
XX Human hepatoma-derived growth factor-2 (HDGF-2) (AAU09404) shows
CC 23% identity and 61% similarity over a 201-amino acid stretch
CC to human HDGF, a mitogen for fibroblasts. Its amino acid sequence
CC was deduced from a cDNA clone (AAU47520) isolated from a human
CC umbilical vein endothelial tissue cDNA library. Recombinant
CC HDGF-2 can be produced in transformed host (e.g. E. coli, Sf9, COS)
CC cells. It is useful in stimulating tissue repair and tissue growth
CC and can also be used to produce antibodies and to screen for
CC antagonists that retard tumour growth and reduce scarring.
XX
SQ Sequence 249 AA;
Query Match 100.0%; Score 1319; DB 18; Length 249;
Best Local Similarity 100.0%; Pred. No. 1.3e-121;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAAVSLRLGDLVWGKLGKRYPPWPGKIYNPPKDLKKPRGKCKFFVKFGTGDHAWIKVEQL 60
DB 1 MAAVSLRLGDLVWGKLGKRYPPWPGKIYNPPKDLKKPRGKCKFFVKFGTGDHAWIKVEQL 60
QY 61 KPYHAHKEEMIKKNGKRFQQAQVDAVEEFLRRAGKDOTSSHNSDDKNNRNSSEERSRP 120
DB 61 KPYHAHKEEMIKKNGKRFQQAQVDAVEEFLRRAGKDOTSSHNSDDKNNRNSSEERSRP 120
QY 121 NSGDEKRKLSLSEGVKKNMGEGKRVSSGSSERGSKSPKRAQSQSPKRGPRPKDEKD 180
DB 121 NSGDEKRKLSLSEGVKKNMGEGKRVSSGSSERGSKSPKRAQSQSPKRGPRPKDEKD 180
QY 181 LTIPESSTVKGMMAGPMAAFKQWPTASEPVKADDPHFHLLSQTEKPAVCYQAITKKLK 240
DB 181 LTIPESSTVKGMMAGPMAAFKQWPTASEPVKADDPHFHLLSQTEKPAVCYQAITKKLK 240
QY 241 ICEDLLPLR 249
DB 241 ICEDLLPLR 249
RESULT 2
AAU20553
ID AAU20553 standard; Protein; 260 AA.
XX
AC AAU20553;
XX
DT 04-DEC-2001 (first entry)
XX
DE Human secreted protein, Seq ID No 545.
XX
KW Immunomodulatory; human immunodeficiency virus; HIV; anaemia; angina;
KW rheumatoid arthritis; antiarteriosclerotic; cardiant; vascular;
KW cerebroprotective; thrombolytic; antimicrobial; ophthalmological;
KW cytostatic; Alzheimer's disease; Parkinson's disease; human cancer;
KW multiple sclerosis; cancer; hyperproliferative disorder; infection;
KW Gaucher's disease; neurological disease; cerebrovascular disorder;
KW thrombosis; wound healing.
XX
OS Homo sapiens.
XX
PN WO200155326-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01347.
XX
PR 31-JAN-2000; 2000US-0179065.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX

```

```

DR WPI; 2001-451931/48.
XX N-PSDB; AAS33262.
XX
PT New nucleic acids and polypeptides, useful for diagnosing, preventing
PT or treating medical conditions -
XX
PS Claim 11; SEQ ID No 545; 753pp; English.
XX
CC The invention relates to novel isolated nucleic acid molecules (I)
CC encoding human secreted proteins (II). (I) and (II) are used to prevent,
CC treat or ameliorate a medical condition in e.g. humans, mice, rabbits,
CC goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used in
CC the prevention, treatment and diagnosis of diseases associated with
CC inappropriate expression of secreted proteins. (I) and complementary
CC sequences may also be used as DNA probes in diagnostic assays (e.g.
CC polymerase chain reactions (PCR)) to detect and quantitate the presence
CC of similar nucleic acid sequences in samples, and so which patients may
CC be in need of restorative therapy. (II) may also be used as antigens in
CC the production of antibodies and in assays to identify modulators
CC (agonists and antagonists) of the expression and activity of the secreted
CC proteins. The anti-(II) antibodies and antagonists may also be used to
CC down regulate expression and activity of (II). The anti-(II) antibodies
CC may also be used as diagnostic agents for detecting the presence of (II)
CC in samples (e.g. by enzyme linked immunosorbant assay (ELISA)). The
CC disorders include for example: immune/autoimmune diseases (e.g. HIV
CC (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis
CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
CC melanomas, neoplasms of the breast or liver, Sezary syndrome and
CC Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,
CC Parkinson's disease and Charcot-Marie-Tooth disease), cardio-/
CC cerebrovascular disorders (e.g. cardiac arrest, tachycardia,
CC angina and thrombosis), infections caused by bacteria, viruses and
CC fungi and ocular disorders (e.g. corneal infections). (I) and (II),
CC agonists, antagonists and antibodies can also be used to promote wound
CC healing, maintain organs before transplantation, and support cell culture
CC of primary tissues. AAU20342-AAU20666 represent human secreted protein
CC amino acid sequences, and related sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification but was obtained in electronic format directly from WIPO
CC at: ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 260 AA;
Query Match 100.0%; Score 1319; DB 22; Length 260;
Best Local Similarity 100.0%; Pred. No. 1.4e-121;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAAVSLRLGDLVWGKLGKRYPPWPGKIYNPPKDLKKPRGKCKFFVKFGTGDHAWIKVEQL 60
DB 12 MAAVSLRLGDLVWGKLGKRYPPWPGKIYNPPKDLKKPRGKCKFFVKFGTGDHAWIKVEQL 71
QY 61 KPYHAHKEEMIKKNGKRFQQAQVDAVEEFLRRAGKDOTSSHNSDDKNNRNSSEERSRP 120
DB 72 KPYHAHKEEMIKKNGKRFQQAQVDAVEEFLRRAGKDOTSSHNSDDKNNRNSSEERSRP 131
QY 121 NSGDEKRKLSLSEGVKKNMGEGKRVSSGSSERGSKSPKRAQSQSPKRGPRPKDEKD 180
DB 132 NSGDEKRKLSLSEGVKKNMGEGKRVSSGSSERGSKSPKRAQSQSPKRGPRPKDEKD 191
QY 181 LTIPESSTVKGMMAGPMAAFKQWPTASEPVKADDPHFHLLSQTEKPAVCYQAITKKLK 240
DB 192 LTIPESSTVKGMMAGPMAAFKQWPTASEPVKADDPHFHLLSQTEKPAVCYQAITKKLK 251
QY 241 ICEDLLPLR 249
DB 252 ICEDLLPLR 260
RESULT 3
AAU21747
ID AAU21747 standard; Protein; 260 AA.
XX
AC AAU21747;

```

XX 04-DEC-2001 (first entry)
XX Novel human neoplastic disease associated polypeptide #180.
XX
XX Human; neoplastic disease associated polypeptide; cancer;
XX hyperproliferative disorder; neural disorder; immune system disorder;
XX muscular disorder; reproductive disorder; gastrointestinal disorder;
XX pulmonary disorder; cardiovascular disorder; renal disorder;
XX neuroprotective; cytostatic; anti inflammatory; vasotrophic.
XX
XX Homo sapiens.
XX
XX WO20015163-A1.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01358.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217496.
XX 14-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225266.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225447.
XX 14-AUG-2000; 2000US-0225757.
XX 14-AUG-2000; 2000US-0225758.
XX 14-AUG-2000; 2000US-0225759.
XX 18-AUG-2000; 2000US-0226279.
XX 22-AUG-2000; 2000US-0226681.
XX 22-AUG-2000; 2000US-0226868.
XX 22-AUG-2000; 2000US-0227182.
XX 23-AUG-2000; 2000US-0227009.
XX 30-AUG-2000; 2000US-0228924.
XX 01-SEP-2000; 2000US-0229287.
XX 01-SEP-2000; 2000US-0229343.
XX 01-SEP-2000; 2000US-0229344.
XX 01-SEP-2000; 2000US-0229345.
XX 05-SEP-2000; 2000US-0229509.
XX 05-SEP-2000; 2000US-0229513.
XX 06-SEP-2000; 2000US-0230437.
XX 06-SEP-2000; 2000US-0230438.
XX 08-SEP-2000; 2000US-0231242.
XX 08-SEP-2000; 2000US-0231243.
XX 08-SEP-2000; 2000US-0231244.
XX 08-SEP-2000; 2000US-0231413.
XX 08-SEP-2000; 2000US-0231414.
XX 08-SEP-2000; 2000US-0232080.
XX 08-SEP-2000; 2000US-0232081.
XX 12-SEP-2000; 2000US-0231968.
XX 14-SEP-2000; 2000US-0232397.
XX 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0241826.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250360.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.

XX AAW69240;
 AC 21-OCT-1998 (first entry)
 DT
 XX
 DE Clone AQ73_3 protein sequence.
 XX
 KW Secreted protein; nutritional source; cell proliferation activity;
 KW cell differentiation activity; immune stimulant; tissue growth activator;
 KW haematopoiesis regulator; anti-inflammatory; tumour invasion suppressor;
 KW tumour inhibitor; clone AQ73_3.
 XX
 OS Homo sapiens.
 PN WO9825962-A2.
 XX
 PD 18-JUN-1998.
 XX
 PF 12-DEC-1997; 97WO-US23224.
 XX
 PR 11-DEC-1997; 97US-0989232.
 PR 13-DEC-1996; 96US-0766263.
 XX
 PA (GEMY) GENETICS INST INC.
 XX
 PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
 PI Racie LA, Spaulding V, Treacy M;
 XX
 DR WPI; 1998-362424/31.
 DR N-PSDB; AAW44859.
 XX
 PT New isolated polynucleotides - obtained from human adult testis,
 PT human adult ovary, human adult brain and human adult heart cDNA
 PT libraries
 XX
 PS Claim 15; Page 66-67; 108pp; English.
 XX
 CC This sequence is secreted protein, encoded by a polynucleotide of the
 CC invention. The DNA was isolated from a human adult ovary cDNA library,
 CC and is designated clone AQ73_3. The DNA sequences and encoded
 CC polypeptides can be used as nutritional sources or supplements, or may
 CC exhibit e.g. cytokine and cell proliferation/differentiation activity,
 CC immune stimulating or suppressing activity, haematopoiesis regulating
 CC activity, receptor/ligand activity, anti-inflammatory activity,
 CC activin/inhibin activity, chemostatic/chemokinetic activity,
 CC cadherin/tumour invasion suppressor activity, tissue growth activity,
 CC tumour inhibition activity or other activities.
 XX
 SQ Sequence 547 AA;
 Query Match 97.9%; Score 1291; DB 19; Length 547;
 Best Local Similarity 99.6%; Pred. No. 2.2e-118;
 Matches 243; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MAAVSLRLGLDVLWGKLGYPWPWPKIYNPPKDLKKPRGKCKFFVFKFGTEHDHAWIKVEQL 60
 DB 1 MAAVSLRLGLDVLWGKLGYPWPWPKIYNPPKDLKKPRGKCKFFVFKFGTEHDHAWIKVEQL 60
 QY 61 KPYHAHKEEMIKINKGRFQQAQVDAVEFLRAKGDQTSNHNSSDDKNRNSSEERSRP 120
 DB 61 KPYHAHKEEMIKINKGRFQQAQVDAVEFLRAKGDQTSNHNSSDDKNRNSSEERSRP 120
 QY 121 NSGDEKRLKLSLSEGVKKNMGEGKKRVSSGSSERGSKSPKRAQEQSPRKRGRPPKDEK 180
 DB 121 NSGDEKRLKLSLSEGVKKNMGEGKKRVSSGSSERGSKSPKRAQEQSPRKRGRPPKDEK 180
 QY 181 LTIPESSTVKGMMAGPMAAFKWKQPTASEPVPKQADPHFHFLLSQTEKPAVCYQAITTKLK 240
 DB 181 LTIPESSTVKGMMAGPMAAFKWKQPTASEPVPKQADPHFHFLLSQTEKPAVCYQAITTKLK 240
 QY 241 ICED 244
 DB 241 ICEE 244

RESULT 6
 AAB90720
 ID AAB90720 standard; Protein; 547 AA.
 XX
 AC AAB90720;
 XX
 DT 07-JUN-2001 (first entry)
 XX
 DE Human AQ73_3 protein sequence SEQ ID 137.
 XX
 KW Human; secreted protein; nutrient; cytokine modulator; proliferation;
 KW differentiation; immune system modulator; tissue growth; chemotactic;
 KW haemostatic; thrombolytic; anti-inflammatory; tumour inhibition;
 KW haematopoiesis.
 XX
 OS Homo sapiens.
 PN WO200119988-A1.
 XX
 PD 22-MAR-2001.
 XX
 PF 14-SEP-2000; 2000WO-US25135.
 XX
 PR 17-SEP-1999; 99US-0398829.
 XX
 PA (GEMY) GENETICS INST INC.
 XX
 PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
 PI Merberg D, Treacy M, Bowman MR, Spaulding V, Agostino MJ;
 XX
 DR WPI; 2001-244801/25.
 DR N-PSDB; AAF98456.
 XX
 PT Isolated nucleic acids encoding polypeptides, useful for modulating
 PT e.g. cytokine and cell proliferation/differentiation activity, the
 PT immune system and hematopoiesis regulating activity -
 XX
 PS Disclosure; Page 468-470; 557pp; English.
 XX
 CC Human cDNA clones represented in AAF98374 - AAF98489 encode secreted
 CC proteins AAB90667 - AAB90750. The cDNA clones are isolated from various
 CC tissue types, and may be used in the prevention, treatment and diagnosis
 CC of diseases associated with inappropriate protein expression. The
 CC polypeptides and nucleic acids may be used as nutrients or to modulate
 CC cytokine and cell proliferation/differentiation activity and may also be
 CC involved in modulation of the immune system. The cDNA sequences,
 CC proteins, their agonists and/or antagonists exhibit haematopoiesis
 CC regulating activity; tissue growth activity; activin/inhibin activity;
 CC chemotactic/chemokinetic activity; haemostatic and thrombolytic
 CC activity; receptor/ligand activity; anti-inflammatory activity;
 CC haematopoiesis activity; cadherin/tumour suppressor activity; and/or
 CC tumour inhibition activity. Included in the invention are probes
 CC represented in AAF98490 - AAF98572 which are specific for the cDNA clones
 CC encoding the secreted proteins.
 XX
 SQ Sequence 547 AA;
 Query Match 97.9%; Score 1291; DB 22; Length 547;
 Best Local Similarity 99.6%; Pred. No. 2.2e-118;
 Matches 243; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MAAVSLRLGLDVLWGKLGYPWPWPKIYNPPKDLKKPRGKCKFFVFKFGTEHDHAWIKVEQL 60
 DB 1 MAAVSLRLGLDVLWGKLGYPWPWPKIYNPPKDLKKPRGKCKFFVFKFGTEHDHAWIKVEQL 60
 QY 61 KPYHAHKEEMIKINKGRFQQAQVDAVEFLRAKGDQTSNHNSSDDKNRNSSEERSRP 120
 DB 61 KPYHAHKEEMIKINKGRFQQAQVDAVEFLRAKGDQTSNHNSSDDKNRNSSEERSRP 120
 QY 121 NSGDEKRLKLSLSEGVKKNMGEGKKRVSSGSSERGSKSPKRAQEQSPRKRGRPPKDEK 180

Db 121 NSGDEKRLSLSEGVKKNMGEGKRVSSGSSRGSKSPKRAEQSPKRGPPKDEKD 180
 Qy 181 LTIPESTVKGMMAGPMAAFKWOPTASEPVKDADPHFHHLLSQTEKPAVCYQAITKKLK 240
 Db 181 LTIPESTVKGMMAGPMAAFKWOPTASEPVKDADPHFHHLLSQTEKPAVCYQAITKKLK 240
 Qy 241 ICED 244
 Db 241 ICEE 244

RESULT 7
 AAO16415
 ID AAO16415 standard; Protein; 553 AA.
 AC AAO16415;
 XX
 DT 10-APR-2003 (first entry)
 XX Human nucleic acid-associated protein (NAAP) - SEQ ID No 12.
 XX Human; nucleic acid-associated protein; NAAP; arteriosclerosis;
 KW cell proliferative disorder; atherosclerosis; cirrhosis; hepatitis; AIDS;
 KW cancer; developmental disorder; renal tubular acidosis; anaemia; asthma;
 KW mental retardation; neurological disorder; Alzheimer's disease; epilepsy;
 KW Parkinson's disease; autoimmune disorder; inflammatory disorder; allergy;
 KW Crohn's disease; transgenic animal; animal model.
 XX
 OS Homo sapiens.
 XX
 PN WO2003000864-A2.
 XX
 PD 03-JAN-2003.
 XX
 PF 20-JUN-2002; 2002WO-US21179.
 XX
 PR 22-JUN-2001; 2001US-300518P.
 PR 29-JUN-2001; 2001US-301787P.
 PR 29-JUN-2001; 2001US-301792P.
 PR 29-JUN-2001; 2001US-301892P.
 PR 29-JUN-2001; 2001US-301893P.
 PR 06-JUL-2001; 2001US-303405P.
 PR 06-JUL-2001; 2001US-303442P.
 PR 15-MAR-2002; 2002US-364438P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Gandhi AR, Swarnakar A, Hafalia AJA, Warren BA, Emerling BM;
 PI Arvizu CS, Ison CH, Honchell CD, Lee EA, Yue H, Forsythe IJ;
 PI Ramkumar J, Griffin JA, Yang J, Sanjanwala MM, Baughn MR;
 PI Borowsky ML, Yao MG, Walla NK, Bandman O, Lal PG, Becha SD;
 PI Lee SY, Richardson TW, Elliott VS, Luo W, Tang YT, Zebarjadian Y;
 PI Lu Y;
 XX
 DR WPI; 2003-201420/19.
 DR N-PSDB; AAL51565.
 XX
 PT New nucleic acid-associated proteins and polynucleotides, useful for
 PT diagnosing, treating or preventing cell proliferative (e.g. cancer),
 PT neurological (e.g. epilepsy or Parkinson's disease), or autoimmune
 PT disorders (e.g. AIDS)
 XX
 PS Claim 1; Page 227-228; 312pp; English.
 XX
 CC The invention comprises the amino acid and coding sequences of human
 CC nucleic acid-associated proteins (NAAP). The DNA and protein sequences of
 CC the invention are useful for diagnosing, treating or preventing disorders
 CC associated with aberrant expression of NAAP, such as: cell proliferative
 CC disorders (e.g. arteriosclerosis, atherosclerosis, cirrhosis, hepatitis
 CC or cancer); developmental disorders (e.g. renal tubular acidosis, anaemia
 CC or mental retardation); neurological disorders (e.g. Alzheimer's disease,
 CC Parkinson's disease or epilepsy); and autoimmune/inflammatory disorders
 CC (e.g. AIDS, allergies, asthma or Crohn's disease). The DNA sequences of

CC the invention are useful for creating transgenic animals to model human
 CC disease. The present amino acid sequence represents a human nucleic acid-
 CC associated protein of the invention.
 XX
 SQ Sequence 553 AA;
 Query Match 97.9%; Score 1291; DB 24; Length 553;
 Best Local Similarity 99.6%; Pred. No. 2.2e-118;
 Matches 243; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 MAAVSLRLGDLVWGKLGYPWPFGKIVNPPKDLKPRGKCFVVKFPGTGDHAWIKVEQL 60
 Db 1 MAAVSLRLGDLVWGKLGYPWPFGKIVNPPKDLKPRGKCFVVKFPGTGDHAWIKVEQL 60
 Qy 61 KPYHAHKEEMIKINKGRFQQAQVDAVEEFILRAKGDQTSNNSSDDKNRRNSSEERSRP 120
 Db 61 KPYHAHKEEMIKINKGRFQQAQVDAVEEFILRAKGDQTSNNSSDDKNRRNSSEERSRP 120
 Qy 121 NSGDEKRLSLSEGVKKNMGEGKRVSSGSSRGSKSPKRAEQSPKRGPPKDEKD 180
 Db 121 NSGDEKRLSLSEGVKKNMGEGKRVSSGSSRGSKSPKRAEQSPKRGPPKDEKD 180
 Qy 181 LTIPESTVKGMMAGPMAAFKWOPTASEPVKDADPHFHHLLSQTEKPAVCYQAITKKLK 240
 Db 181 LTIPESTVKGMMAGPMAAFKWOPTASEPVKDADPHFHHLLSQTEKPAVCYQAITKKLK 240
 Qy 241 ICED 244
 Db 241 ICEE 244

RESULT 8
 AAU21690
 ID AAU21690 standard; Protein; 550 AA.
 XX
 AC AAU21690;
 XX
 DT 04-DEC-2001 (first entry)
 XX
 DE Novel human neoplastic disease associated polypeptide #123.
 XX
 KW Human; neoplastic disease associated polypeptide; cancer;
 KW hyperproliferative disorder; neural disorder; immune system disorder;
 KW muscular disorder; reproductive disorder; gastrointestinal disorder;
 KW pulmonary disorder; cardiovascular disorder; renal disorder;
 KW neuroprotective; cytostatic; anti-inflammatory; vasotropic.
 XX
 OS Homo sapiens.
 XX
 PN WO200155163-A1.
 XX
 PD 02-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US01358.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184654.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.

PR	14-AUG-2000;	2000US-022451100;	2000US-02245119
PR	14-AUG-2000;	2000US-022451200;	2000US-022451213
PR	14-AUG-2000;	2000US-022451210;	2000US-022451219
PR	14-AUG-2000;	2000US-022451214;	2000US-022451214
PR	14-AUG-2000;	2000US-022451266;	2000US-022451266
PR	14-AUG-2000;	2000US-022451267;	2000US-022451267
PR	14-AUG-2000;	2000US-022451268;	2000US-022451268
PR	14-AUG-2000;	2000US-022527000;	2000US-022527000
PR	14-AUG-2000;	2000US-022527070;	2000US-022527070
PR	14-AUG-2000;	2000US-022544447;	2000US-022544447
PR	14-AUG-2000;	2000US-022575757;	2000US-022575757
PR	14-AUG-2000;	2000US-022575758;	2000US-022575758
PR	14-AUG-2000;	2000US-022575759;	2000US-022575759
PR	18-AUG-2000;	2000US-022627779;	2000US-022627779
PR	22-AUG-2000;	2000US-022666811;	2000US-022666811
PR	22-AUG-2000;	2000US-022668686;	2000US-022668686
PR	23-AUG-2000;	2000US-022711821;	2000US-022711821
PR	23-AUG-2000;	2000US-022720009;	2000US-022720009
PR	30-AUG-2000;	2000US-022892924;	2000US-022892924
PR	01-SEP-2000;	2000US-022929343;	2000US-022929343
PR	01-SEP-2000;	2000US-022931444;	2000US-022931444
PR	01-SEP-2000;	2000US-022931445;	2000US-022931445
PR	05-SEP-2000;	2000US-022959009;	2000US-022959009
PR	06-SEP-2000;	2000US-023045133;	2000US-023045133
PR	06-SEP-2000;	2000US-023043737;	2000US-023043737
PR	06-SEP-2000;	2000US-023043838;	2000US-023043838
PR	08-SEP-2000;	2000US-023121242;	2000US-023121242
PR	08-SEP-2000;	2000US-023121243;	2000US-023121243
PR	08-SEP-2000;	2000US-023121444;	2000US-023121444
PR	08-SEP-2000;	2000US-023141414;	2000US-023141414
PR	08-SEP-2000;	2000US-023140800;	2000US-023140800
PR	12-SEP-2000;	2000US-023108101;	2000US-023108101
PR	12-SEP-2000;	2000US-023196868;	2000US-023196868
PR	14-SEP-2000;	2000US-023233977;	2000US-023233977
PR	14-SEP-2000;	2000US-023233998;	2000US-023233998
PR	14-SEP-2000;	2000US-023233999;	2000US-023233999
PR	14-SEP-2000;	2000US-023234000;	2000US-023234000
PR	14-SEP-2000;	2000US-023306303;	2000US-023306303
PR	14-SEP-2000;	2000US-023306304;	2000US-023306304
PR	14-SEP-2000;	2000US-023306305;	2000US-023306305
PR	21-SEP-2000;	2000US-023424223;	2000US-023424223
PR	21-SEP-2000;	2000US-023424274;	2000US-023424274
PR	25-SEP-2000;	2000US-023499797;	2000US-023499797
PR	25-SEP-2000;	2000US-023499898;	2000US-023499898
PR	26-SEP-2000;	2000US-023494844;	2000US-023494844
PR	27-SEP-2000;	2000US-023568334;	2000US-023568334
PR	27-SEP-2000;	2000US-023568336;	2000US-023568336
PR	29-SEP-2000;	2000US-023636727;	2000US-023636727
PR	29-SEP-2000;	2000US-023636728;	2000US-023636728
PR	02-OCT-2000;	2000US-023680202;	2000US-023680202
PR	02-OCT-2000;	2000US-023700377;	2000US-023700377
PR	02-OCT-2000;	2000US-023700388;	2000US-023700388
PR	02-OCT-2000;	2000US-023700390;	2000US-023700390
PR	20-OCT-2000;	2000US-024109260;	2000US-024109260
PR	20-OCT-2000;	2000US-024122622;	2000US-024122622
PR	20-OCT-2000;	2000US-024178555;	2000US-024178555
PR	20-OCT-2000;	2000US-024178666;	2000US-024178666
PR	20-OCT-2000;	2000US-024178777;	2000US-024178777
PR	13-OCT-2000;	2000US-023999335;	2000US-023999335
PR	13-OCT-2000;	2000US-023999337;	2000US-023999337
PR	20-OCT-2000;	2000US-024180808;	2000US-024180808
PR	20-OCT-2000;	2000US-024180909;	2000US-024180909
PR	20-OCT-2000;	2000US-024192622;	2000US-024192622
PR	20-OCT-2000;	2000US-024122622;	2000US-024122622

PR	08-NOV-2000;	2000US-02464377;
PR	08-NOV-2000;	2000US-0246478;
PR	08-NOV-2000;	2000US-0246523;
PR	08-NOV-2000;	2000US-0246524;
PR	08-NOV-2000;	2000US-0246525;
PR	08-NOV-2000;	2000US-0246526;
PR	08-NOV-2000;	2000US-0246527;
PR	08-NOV-2000;	2000US-0246528;
PR	08-NOV-2000;	2000US-0246532;
PR	08-NOV-2000;	2000US-0246609;
PR	08-NOV-2000;	2000US-0246610;
PR	08-NOV-2000;	2000US-0246611;
PR	08-NOV-2000;	2000US-0246617;
PR	08-NOV-2000;	2000US-0246620;
PR	17-NOV-2000;	2000US-0249207;
PR	17-NOV-2000;	2000US-0249208;
PR	17-NOV-2000;	2000US-0249209;
PR	17-NOV-2000;	2000US-0249210;
PR	17-NOV-2000;	2000US-0249211;
PR	17-NOV-2000;	2000US-0249212;
PR	17-NOV-2000;	2000US-0249213;
PR	17-NOV-2000;	2000US-0249214;
PR	17-NOV-2000;	2000US-0249215;
PR	17-NOV-2000;	2000US-0249216;
PR	17-NOV-2000;	2000US-0249217;
PR	17-NOV-2000;	2000US-0249218;
PR	17-NOV-2000;	2000US-0249244;
PR	17-NOV-2000;	2000US-0249245;
PR	17-NOV-2000;	2000US-0249264;
PR	17-NOV-2000;	2000US-0249265;
PR	17-NOV-2000;	2000US-0249297;
PR	17-NOV-2000;	2000US-0249299;
PR	17-NOV-2000;	2000US-0249299;
PR	17-NOV-2000;	2000US-0249300;
PR	01-DEC-2000;	2000US-0250160;
PR	01-DEC-2000;	2000US-0250391;
PR	05-DEC-2000;	2000US-0251030;
PR	05-DEC-2000;	2000US-0251988;
PR	06-DEC-2000;	2000US-0256719;
PR	06-DEC-2000;	2000US-0251479;
PR	08-DEC-2000;	2000US-0251856;
PR	08-DEC-2000;	2000US-0251868;
PR	08-DEC-2000;	2000US-0251869;
PR	08-DEC-2000;	2000US-0251989;
PR	08-DEC-2000;	2000US-0251990;
PR	11-DEC-2000;	2000US-0254097;
PR	05-JAN-2001;	2001US-0259678;
XX		
PA	(HUNA-) HUMAN GENOME SCI INC	

CC neoplastic disease associated polypeptides of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 550 AA;
 Query Match 96.9%; Score 1278; DB 22; Length 550;
 Best Local Similarity 99.6%; Pred. No. 4.2e-117;
 Matches 240; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 VSLRLGDLVWGKLGRIYPPWPGKIVNPPKDLKKPRGKCFVFKFGTGDHAWIKVEQLKPY 63
 DB 7 VSLRLGDLVWGKLGRIYPPWPGKIVNPPKDLKKPRGKCFVFKFGTGDHAWIKVEQLKPY 66
 QY 64 HAHEEMIKINKGRKRFQAVDAVEEFLLRAKGDQTSNHNSSDDKNRNSSEERSRNSG 123
 DB 67 HAHEEMIKINKGRKRFQAVDAVEEFLLRAKGDQTSNHNSSDDKNRNSSEERSRNSG 126
 QY 124 DEKRKLSLSEGVKVMGEGKRVSSGSRGSKSPKRAQEQSPKRGPRPPKDEKDLTI 183
 DB 127 DEKRKLSLSEGVKVMGEGKRVSSGSRGSKSPKRAQEQSPKRGPRPPKDEKDLTI 186
 QY 184 PESTVKGMMAGPMAAFKQPTASEPVKDADPHFHLLSQTEKPAVCYQAITKKLICE 243
 DB 187 PESTVKGMMAGPMAAFKQPTASEPVKDADPHFHLLSQTEKPAVCYQAITKKLICE 246
 QY 244 D 244
 DB 247 E 247

RESULT 9
 AAU20515
 ID AAU20515 standard; Protein; 269 AA.
 XX AC AAU20515;
 XX DT 06-DEC-2001 (first entry)
 DE Human secreted protein, Seq ID No 507.
 XX
 KW Immunomodulatory; human immunodeficiency virus; HIV; anaemia; angina;
 KW rheumatoid arthritis; antiarteriosclerotic; cardiant; vascular;
 KW cerebroprotective; thrombolytic; antimicrobial; ophthalmological;
 KW cytosstatic; Alzheimer's disease; Parkinson's disease; human; cancer;
 KW multiple sclerosis; cancer; hyperproliferative disorder; infection;
 KW Gaucher's disease; neurological disease; cerebrovascular disorder;
 KW thrombosis; wound healing.
 XX
 OS Homo sapiens.
 XX
 PN WO200155326-A2.
 XX
 PD 02-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US01347.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI; 2001-451931/48.
 DR N-PSDB; AAS33224.
 XX
 PT New nucleic acids and polypeptides, useful for diagnosing, preventing
 PT or treating medical conditions -
 XX
 PS Claim 11; SEQ ID No 507; 753bp; English.
 XX
 CC The invention relates to novel isolated nucleic acid molecules (I)

CC encoding human secreted proteins (II). (I) and (II) are used to prevent,
 CC treat or ameliorate a medical condition in e.g. humans, mice, rabbits,
 CC goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used in
 CC the prevention, treatment and diagnosis of diseases associated with
 CC inappropriate expression of secreted proteins. (I) and complementary
 CC sequences may also be used as DNA probes in diagnostic assays (e.g.
 CC polymerase chain reactions (PCR)) to detect and quantify the presence
 CC of similar nucleic acid sequences in samples, and so which patients may
 CC be in need of restorative therapy. (II) may also be used as antigens in
 CC the production of antibodies and in assays to identify modulators
 CC (agonists and antagonists) of the expression and activity of the secreted
 CC proteins. The anti-(II) antibodies and antagonists may also be used to
 CC down regulate expression and activity of (II). The anti-(II) antibodies
 CC may also be used as diagnostic agents for detecting the presence of (II)
 CC in samples (e.g. by enzyme linked immunosorbant assay (ELISA)). The
 CC disorders include for example: immune/autoimmune diseases (e.g. HIV
 CC (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis
 CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
 CC melanomas, neoplasms of the breast or liver, Sezary syndrome and
 CC Parkinson's disease), neurological diseases (e.g. Alzheimer's disease,
 CC cerebrovascular disorders (e.g. cardiac arrest, tachycardia,
 CC angina and thrombosis), infections caused by bacteria, viruses and
 CC fungi and ocular disorders (e.g. corneal infections). (I) and (II),
 CC agonists, antagonists and antibodies can also be used to promote wound
 CC healing, maintain organs before transplantation, and support cell culture
 CC of primary tissues. AAU20342-AAU20666 represent human secreted protein
 CC amino acid sequences, and related sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification but was obtained in electronic format directly from WIPO
 CC at: ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 269 AA;
 Query Match 95.8%; Score 1263; DB 22; Length 269;
 Best Local Similarity 98.8%; Pred. No. 4.7e-116;
 Matches 238; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 VSLRLGDLVWGKLGRIYPPWPGKIVNPPKDLKKPRGKCFVFKFGTGDHAWIKVEQLKPY 63
 DB 1 VSLRLGDLVWGKLGRIYPPWPGKIVNPPKDLKKPRGKCFVFKFGTGDHAWIKVEQLKPY 60
 QY 64 HAHEEMIKINKGRKRFQAVDAVEEFLLRAKGDQTSNHNSSDDKNRNSSEERSRNSG 123
 DB 61 HAHEEMIKINKGRKRFQAVDAVEEFLLRAKGDQTSNHNSSDDKNRNSSEERSRNSG 120
 QY 124 DEKRKLSLSEGVKVMGEGKRVSSGSRGSKSPKRAQEQSPKRGPRPPKDEKDLTI 183
 DB 121 DEKRKLSLSEGVKVMGEGKRVSSGSRGSKSPKRAQEQSPKRGPRPPKDEKDLTI 180
 QY 184 PESTVKGMMAGPMAAFKQPTASEPVKDADPHFHLLSQTEKPAVCYQAITKKLICE 243
 DB 181 PESTVKGMMAGPMAAFKQPTASEPVKDADPHFHLLSQTEKPAVCYQAITKKLICE 240
 QY 244 D 244
 DB 241 E 241

RESULT 10
 AAU21813
 ID AAU21813 standard; Protein; 269 AA.
 XX AC AAU21813;
 XX DT 06-DEC-2001 (first entry)
 XX
 DE Novel human neoplastic disease associated polypeptide #246.
 XX
 KW Human; neoplastic disease associated polypeptide; cancer;
 KW hyperproliferative disorder; neural disorder; immune system disorder;
 KW muscular disorder; reproductive disorder; gastrointestinal disorder;
 KW pulmonary disorder; cardiovascular disorder; renal disorder;

KW neuroprotective; cytostatic; anti inflammatory; vasotropic.

XX Homo sapiens.

PN WO200155163-A1.

XX PD 02-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US01358.

XX PR 31-JAN-2000; 2000US-0179065.

XX PR 04-FEB-2000; 2000US-0180628.

XX PR 24-FEB-2000; 2000US-0184664.

XX PR 02-MAR-2000; 2000US-0186350.

XX PR 16-MAR-2000; 2000US-0189874.

XX PR 17-MAR-2000; 2000US-0190076.

XX PR 18-APR-2000; 2000US-0198123.

XX PR 19-MAY-2000; 2000US-0205515.

XX PR 07-JUN-2000; 2000US-0209467.

XX PR 28-JUN-2000; 2000US-0214886.

XX PR 30-JUN-2000; 2000US-0215135.

XX PR 07-JUL-2000; 2000US-0216647.

XX PR 07-JUL-2000; 2000US-0216880.

XX PR 11-JUL-2000; 2000US-0217487.

XX PR 11-JUL-2000; 2000US-0217496.

XX PR 14-JUL-2000; 2000US-0218290.

XX PR 26-JUL-2000; 2000US-0220963.

XX PR 26-JUL-2000; 2000US-0220964.

XX PR 14-AUG-2000; 2000US-0224518.

XX PR 14-AUG-2000; 2000US-0224519.

XX PR 14-AUG-2000; 2000US-0225213.

XX PR 14-AUG-2000; 2000US-0225214.

XX PR 14-AUG-2000; 2000US-0225266.

XX PR 14-AUG-2000; 2000US-0225267.

XX PR 14-AUG-2000; 2000US-0225268.

XX PR 14-AUG-2000; 2000US-0225270.

XX PR 14-AUG-2000; 2000US-0225447.

XX PR 14-AUG-2000; 2000US-0225757.

XX PR 14-AUG-2000; 2000US-0225758.

XX PR 18-AUG-2000; 2000US-0225759.

XX PR 22-AUG-2000; 2000US-0226279.

XX PR 22-AUG-2000; 2000US-0226681.

XX PR 22-AUG-2000; 2000US-0226868.

XX PR 22-AUG-2000; 2000US-0227182.

XX PR 23-AUG-2000; 2000US-0227009.

XX PR 30-AUG-2000; 2000US-0228924.

XX PR 01-SEP-2000; 2000US-0229287.

XX PR 01-SEP-2000; 2000US-0229343.

XX PR 01-SEP-2000; 2000US-0229344.

XX PR 01-SEP-2000; 2000US-0229345.

XX PR 05-SEP-2000; 2000US-0229509.

XX PR 05-SEP-2000; 2000US-0229513.

XX PR 06-SEP-2000; 2000US-0230437.

XX PR 06-SEP-2000; 2000US-0230438.

XX PR 08-SEP-2000; 2000US-0231242.

XX PR 08-SEP-2000; 2000US-0231243.

XX PR 08-SEP-2000; 2000US-0231244.

XX PR 08-SEP-2000; 2000US-0231413.

XX PR 08-SEP-2000; 2000US-0231414.

XX PR 08-SEP-2000; 2000US-0232080.

XX PR 08-SEP-2000; 2000US-0232081.

XX PR 12-SEP-2000; 2000US-0231968.

XX PR 14-SEP-2000; 2000US-0232397.

XX PR 14-SEP-2000; 2000US-0232398.

XX PR 14-SEP-2000; 2000US-0232399.

XX PR 14-SEP-2000; 2000US-0232400.

XX PR 14-SEP-2000; 2000US-0232401.

XX PR 14-SEP-2000; 2000US-0233063.

XX PR 14-SEP-2000; 2000US-0233064.

XX PR 21-SEP-2000; 2000US-0233065.

XX PR 21-SEP-2000; 2000US-0234223.

XX PR 21-SEP-2000; 2000US-0234274.

XX PR 25-SEP-2000; 2000US-0234997.

PR 25-SEP-2000; 2000US-0234998.

PR 26-SEP-2000; 2000US-0235484.

PR 27-SEP-2000; 2000US-0235834.

PR 27-SEP-2000; 2000US-0235836.

PR 29-SEP-2000; 2000US-0236327.

PR 29-SEP-2000; 2000US-0236367.

PR 29-SEP-2000; 2000US-0236368.

PR 29-SEP-2000; 2000US-0236369.

PR 29-SEP-2000; 2000US-0236370.

PR 02-OCT-2000; 2000US-0236802.

PR 02-OCT-2000; 2000US-0237037.

PR 02-OCT-2000; 2000US-0237038.

PR 02-OCT-2000; 2000US-0237039.

PR 12-OCT-2000; 2000US-0237040.

PR 13-OCT-2000; 2000US-0239935.

PR 13-OCT-2000; 2000US-0239937.

PR 20-OCT-2000; 2000US-0240960.

PR 20-OCT-2000; 2000US-0241221.

PR 20-OCT-2000; 2000US-0241785.

PR 20-OCT-2000; 2000US-0241786.

PR 20-OCT-2000; 2000US-0241787.

PR 20-OCT-2000; 2000US-0241808.

PR 20-OCT-2000; 2000US-0241809.

PR 20-OCT-2000; 2000US-0241826.

PR 01-NOV-2000; 2000US-0244617.

PR 08-NOV-2000; 2000US-0246474.

PR 08-NOV-2000; 2000US-0246475.

PR 08-NOV-2000; 2000US-0246476.

PR 08-NOV-2000; 2000US-0246477.

PR 08-NOV-2000; 2000US-0246478.

PR 08-NOV-2000; 2000US-0246523.

PR 08-NOV-2000; 2000US-0246524.

PR 08-NOV-2000; 2000US-0246525.

PR 08-NOV-2000; 2000US-0246526.

PR 08-NOV-2000; 2000US-0246527.

PR 08-NOV-2000; 2000US-0246528.

PR 08-NOV-2000; 2000US-0246532.

PR 08-NOV-2000; 2000US-0246609.

PR 08-NOV-2000; 2000US-0246610.

PR 08-NOV-2000; 2000US-0246611.

PR 17-NOV-2000; 2000US-0249207.

PR 17-NOV-2000; 2000US-0249208.

PR 17-NOV-2000; 2000US-0249209.

PR 17-NOV-2000; 2000US-0249210.

PR 17-NOV-2000; 2000US-0249211.

PR 17-NOV-2000; 2000US-0249212.

PR 17-NOV-2000; 2000US-0249213.

PR 17-NOV-2000; 2000US-0249214.

PR 17-NOV-2000; 2000US-0249215.

PR 17-NOV-2000; 2000US-0249216.

PR 17-NOV-2000; 2000US-0249217.

PR 17-NOV-2000; 2000US-0249218.

PR 17-NOV-2000; 2000US-0249244.

PR 17-NOV-2000; 2000US-0249245.

PR 17-NOV-2000; 2000US-0249264.

PR 17-NOV-2000; 2000US-0249265.

PR 17-NOV-2000; 2000US-0249297.

PR 17-NOV-2000; 2000US-0249299.

PR 17-NOV-2000; 2000US-0249300.

PR 01-DEC-2000; 2000US-0250160.

PR 01-DEC-2000; 2000US-0250391.

PR 05-DEC-2000; 2000US-0251030.

PR 05-DEC-2000; 2000US-0251988.

PR 05-DEC-2000; 2000US-0256719.

PR 06-DEC-2000; 2000US-0251479.

PR 08-DEC-2000; 2000US-0251856.

PR 08-DEC-2000; 2000US-0251868.

PR 08-DEC-2000; 2000US-0251869.

PR 08-DEC-2000; 2000US-0251989.

PR 08-DEC-2000; 2000US-0251990.

PR 11-DEC-2000; 2000US-0254097.

PR 05-JAN-2001; 2001US-0259678.

```
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Barash SC, Ruben SM;
XX XX
XX XX
XX WPI; 2001-465558/50.
XX DR N-PSDB; AAS35012.
XX XX
XX Novel polypeptides and polynucleotides useful as diagnostic reagents to
XX PT diagnose diseases or disorders associated with aberrant expression or
XX PT activity of polypeptides, and for treating cancers, rheumatoid
XX PT arthritis
XX XX
XX Claim 11; SEQ ID NO 540; 687bp; English.
XX PS
XX XX
XX The present invention relates to the isolation of novel human neoplastic
XX CC disease associated polypeptides, and cDNA (AAS34767-AAS35050) and DNA
XX CC sequences encoding for these polypeptides. The sequences of the
XX CC invention are useful in the diagnosis, treatment, prevention and/or
XX CC prognosis of disorders involving neoplastic disease such as
XX CC hyperproliferative disorders (e.g. leukemia, bone cancer, bladder
XX CC cancer, brain stem glioma, adult liver cancer, childhood cerebellar
XX CC astrocytoma, or Hodgkin's lymphoma). The sequences of the invention may
XX CC also be useful for treating other disorders such as neural disorders,
XX CC immune system disorders, muscular disorders, reproductive disorders,
XX CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
XX CC and renal disorders. The polynucleotide sequences of the invention are
XX CC also useful in gene therapy. AAU21568-AAU21851 represent the novel human
XX CC neoplastic disease associated polypeptides of the invention.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX XX
XX Sequence 269 AA;
XX
Query Match 95.8%; Score 1263; DB 22; Length 269;
Best Local Similarity 98.8%; Pred. NO. 4.7e-116;
Matches 238; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 4 VSLRLGLVWGKLGRLYPWPVKIYNPPKDLKKPRGKCFVFKFGTEDHAMIKVEQLKP 63
Db 1 VSLRLGLVWGKLGRLYPWPVKIYNPPKDLKKPRGKCFVFKFGTEDHAMIKVEQLKP 60
Qy 64 HAHKEEMIKNKGRFQOAVDAVEFLRAKQDQTSNHSDDKNRNSSEERSRPN 123
Db 61 HAHKEEMIKNKGRFQOAVDAVEFLRAKQDQTSNHSDDKNRNSSEERSRPN 120
Qy 124 DEKRLSLSEGKVKKNMGEGKRVSSGSSGSKSPKRAQEQSPRKRGRPPKDEKDLTI 183
Db 121 DEKRLSLSEGKVKKNMGEGKRVSSGSSGSKSPKRAQEQSKRKRGRPPKDEKDLTI 180
Qy 184 PESSTVKGMMAGPMAAFKQWPTASEPVKDADPHFHLFSQTEPAVCYQAITKKLICE 243
Db 181 PESSTVKGMMAGPMAAFKQWPTASEPVKDADPHFHLFSQTEPAVCYQAITKKLICE 240
Qy 244 D 244
Db 241 E 241
RESULT 11
ABB58603
ID ABB58603 standard; Protein; 475 AA.
XX
XX ABB58603;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 2601.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical.
XX
XX
```

```
OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX XX
XX PD 27-SEP-2001.
XX XX
XX 23-MAR-2001; 2001WO-US09231.
XX XX
XX 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX XX
XX (PEKE ) PE CORP NY.
XX PA
XX XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX PI WPI; 2001-656860/75.
XX DR N-PSDB; ABL02706.
XX XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions
XX XX
XX Disclosure; SEQ ID NO 2601; 21pp + Sequence Listing; English.
XX PS
XX The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins
XX CC (ABBS7737-ABBS72072).
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX XX
XX Sequence 475 AA;
XX
Query Match 14.6%; Score 193; DB 22; Length 475;
Best Local Similarity 27.7%; Pred. No. 3.2e-10;
Matches 64; Conservative 32; Mismatches 89; Indels 46; Gaps 8;
Qy 2 AAVSLRLGLVWGKLGRLYPWPVKIYNPPKDLKKPRGKCFVFKFGTEDHAMIKVEQLK 61
Db 6 AAASYSIGDLVFAKVGKYPWPVKIT-----KSNNNKYNVYFYTGTANIKLEDLF 58
Qy 62 PYHAHKEEMI--KINKGRFQOAVDAVEFLR-----RAKQDQTSNHSDDKNRR 111
Db 59 PYASNKERFATEKIMKRAKFIETADQIESALRGEDSAPIDLDDGAEPVAPTTGD----- 113
Qy 112 NSSEERSRPNSSGDEKRLSLSEGKVKKNMGEGKRVSS-----GSSERGSKSPKRAQE 165
Db 114 GVTEEPKPEPGPEPATAAPVAAAEEKKSGTKTKTKAPPRHVDGDSAGAE----- 166
Qy 166 QSPKRGRPKPKDKDLTIPESSSTVKGMMAGPMAAFKQWPTAS-EPVKDADP 215
Db 167 -----APPAKERVETEGLATATAAIPAAPAAA---TPTSSKSKVKKSKP 207
RESULT 12
ABB72327
ID ABB72327 standard; Protein; 202 AA.
XX
XX ABB72327;
XX AC
XX 04-APR-2002 (first entry)
XX
XX Rat protein isolated from skin cells SEQ ID NO: 651.
XX
XX Human; rat; mouse; skin cell; skin wound; cancer; growth defect;
XX KW developmental defect; inflammatory disease; dermatological; vulnary;
XX KW immunomodulator; anti-inflammatory; cytostatic; neuroprotective.
XX
```

OS Rattus sp.
XX WO200190357-A1.
XX 29-NOV-2001.
XX 24-MAY-2001; 2001WO-NZ00099.
XX 24-MAY-2000; 2000US-206650P.
PR 25-JUL-2000; 2000US-221232P.
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX Watson JD, Strachan L, Sleeman M, Orrust R, Murison JG, Kumble KD;
PI WPI: 2002-122020/16.
XX N-PSDB; ABL35012.
XX New polynucleotides and polypeptides encoded by the polynucleotides
PT isolated from skin cells, useful for treating skin wounds, cancers,
PT growth and developmental defects, inflammatory diseases, or for
PT modulating immune responses -
XX Claim 4; Page 408-409; 466pp; English.
XX The present invention provides the protein and coding sequences of cDNAs
CC isolated from human, murine and rat skin cell libraries. The sequences
CC can be used in the development of therapeutic agents useful in the
CC treatment of skin diseases, including skin wounds, cancer, growth
CC defects, developmental defects and inflammatory diseases. The proteins
CC have important roles in the induction of hair growth, cell proliferation
CC and cell-cell interaction, in maintaining tissue integrity, in wound
CC healing and in modulating immune responses. The present sequence is a
CC polypeptide of the invention.
XX Sequence 202 AA;
SQ Query Match 14.4%; Score 190.5; DB 23; Length 202;
Best Local Similarity 28.7%; Pred. No. 1.7e-10;
Matches 56; Conservative 37; Mismatches 77; Indels 25; Gaps 6;
QY 7 RIGDLVMGKLGYPWPWGKINVPNDLKKPRGKCFVFPCTEDHAWIKVEQLKPYAH 66
Db 10 KAGDLVFAKMGYPHWPARIDELPEGAVPPANK-YPIFFGTHTAFLGPKDLFPYKEY 68
QY 67 KEEMIKINKKRFQQAQVAVE-----EFLRAKGDQDTSSHNSDDKRRNSSEERSRN 121
Db 69 KDKFGSKNRKRGFNGLWEIENNPVGKFTGYQTIQQSSSETEGEGGNTADASSE---E 124
QY 122 SGDEKRLSLSEGVKKNMGEGKRVSGSSERSGSKPLKRAEQSPRRGRPPPKDEKDL 181
Db 125 EGD-----RVEDGKGRKNEKGGSKRKSYTSKSSKQSRKSPG--DEDDKDC 170
QY 182 TTPES-STVKGMMAG 195
Db 171 KEENKSSSEGGDAG 185
RESULT 13
AAAY28502
ID AAY28502 standard; Protein; 203 AA.
XX AC AAY28502;
XX DT 19-OCT-1999 (first entry)
XX HGPH2 Human Growth Factor Homologue 2.
XX HGPH2; human growth factor homologue 2; cell proliferation; HDGF;
KW hepatoma derived growth factor; protein kinase C; casein kinase II;
KW glycosylation site; CAMP dependent protein kinase; cancer; hepatitis;
XX anaemia.

OS Homo sapiens.
XX Key Location/Qualifiers
XX Modified-site 75
FT /note= "Potential protein kinase C phosphorylation"
FT Modified-site 106
FT /note= "Potential casein kinase II phosphorylation site"
FT Modified-site 108
FT /note= "Potential casein kinase II phosphorylation site"
FT Modified-site 110
FT /note= "Potential casein kinase II phosphorylation site"
FT Modified-site 121
FT /note= "Potential casein kinase II phosphorylation site"
FT Modified-site 122
FT /note= "Potential casein kinase II phosphorylation site"
FT Modified-site 144
FT /note= "Potential protein kinase C phosphorylation site"
FT Modified-site 149
FT /note= "Potential CAMP- and cGMP-dependent protein
kinase phosphorylation site"
FT Modified-site 151
FT /note= "Potential protein kinase C phosphorylation site"
FT Modified-site 152
FT /note= "Potential protein kinase C phosphorylation site"
FT Modified-site 155
FT /note= "Potential protein kinase C phosphorylation site"
FT Modified-site 156
FT /note= "Potential CAMP- and cGMP-dependent protein
kinase phosphorylation site"
FT Modified-site 159
FT /note= "Potential protein kinase C phosphorylation site"
FT Modified-site 162
FT /note= "Potential casein kinase II phosphorylation site"
FT Modified-site 176
FT /note= "Potential N-glycosylation site"
FT Modified-site 178
FT /note= "Potential casein kinase II phosphorylation site"
FT Modified-site 187
FT /note= "Potential N-glycosylation site"
FT Modified-site 191
FT /note= "Potential N-glycosylation site"
FT Modified-site 192
FT /note= "Potential casein kinase II phosphorylation site"
XX WO9938976-A2.
XX 05-AUG-1999.
XX 12-JAN-1999; 99WO-US00654.
XX 29-JAN-1998; 98US-0015412.
XX (INCY-) INCYTE PHARM INC.
XX Baughn M, Corley NC, Guegler KJ, Hillman JL, Lal P;
PI Tang YT;
XX WPI: 1999-469331/39.
XX N-PSDB; AAZ00048.
XX New human growth factor homologs and their use in the diagnosis,
PT treatment and prevention of cell proliferative and developmental
PT disorders
XX Claim 1; Fig 2; 88pp; English.
XX This sequence is the HGPH2 Human Growth Factor Homologue 2. The HGPH2
CC protein has structural and chemical homology to the human hepatoma
CC derived growth factor (HDGF). HGPH2 is 203 amino acids in length
CC and has three potential N-glycosylation sites, two potential CAMP- and
CC cGMP-dependent protein kinase phosphorylation sites, eight potential
CC casein kinase II phosphorylation sites and six potential protein kinase C
CC phosphorylation sites. HGPHs are expressed in cancerous tissues and

CC appear to play a role in cell proliferative and developmental disorders.
 CC A purified antagonist of HGFH can be administered to a subject as a
 CC method of treating or preventing a cell proliferative disorder e.g.
 CC atherosclerosis, hepatitis, and some cancers. Antibodies against HGFH and
 CC antisense sequences may also be used as antagonist for treating the
 CC above. Developmental disorders treated by HGFH include renal tubular
 CC acidosiis, anaemia, Cushing's syndrome, achondroplastic dwarfism,
 CC Duchenne and Becker muscular dystrophy, and epilepsy. HGFH antibodies and
 CC HGFH polynucleotides and polypeptides may also be used in various
 CC diagnostic methods.

SQ Sequence 203 AA;
 Query Match 14.3%; Score 189; DB 20; Length 203;
 Best Local Similarity 28.2%; Pred. No. 2.5e-10;
 Matches 55; Conservative 37; Mismatches 79; Indels 24; Gaps 5;
 QY 7 RLGLDVLWGKLGKRYPPWPGKIVNPPKDLKPRGKCKFFVKGTEGTHAWIKVEQLKPYHAAH 66
 DB 10 KAGDLVFAKMKGYPHWPARIDELPEGAVKPPANK-YPIFFFGTHETAFLGPKOLFPPYKEY 68
 QY 67 KEEMIKINKGRFOQAQDAVE-----EFLRRAGKQDOTSHNSDDKNRNSSEERSRPN 121
 DB 69 KDKFGKSNKRKGFNEGLWEIENNPVGFYQAIQQOSSSETEGEGGNTADASSE----- 123
 QY 122 SGDEKRLKLSISEGKVKKNMGEGKRVSSGSSERGSKSPKRAQEQSPRKRGRPPKDEKOL 181
 DB 124 -----EEGDRVEEDGKGRKNEKAGSKRKSYTSKKSKQSRKSPG--DEDDKDC 171
 QY 182 TTPES-STVKGMMAG 195
 DB 172 KEENKSSSEGGDAG 186

RESULT 14
 AAB13521
 ID AAB13521 standard; Protein; 203 AA.
 XX AC AAB13521;
 XX AC AAB13521;

DT 13-NOV-2000 (first entry)

XX Human hepatoma-derived growth factor homologous polypeptide.

XX Human; hepatoma-derived growth factor homologous polypeptide;
 KW huHGFh; cell proliferation; haematopoiesis; angiogenesis;
 KW lymphocyte proliferation; infection; autoimmune disease;
 KW vascular disease; cancer.

XX Homo sapiens.

OS WO200037492-A2.

PN 29-JUN-2000.

XX 22-DEC-1999; 99WO-US30932.

XX 22-DEC-1998; 98US-0113344.

XX (ELIL) LILLY & CO ELI.

XX Na S;

XX WPI; 2000-442638/38.

DR N-PSDB; AAA65004.

XX Novel human homologues of hepatoma-derived growth factor homologous
 PT nucleic acids, polypeptides useful as probes or amplification primers
 PT in the detection, quantitation or isolation of gene sequences or
 PT transcripts

XX Claim 9; Page 74-75; 75pp; English.

CC The present sequence is the human homologue of the hepatoma-derived
 CC growth factor (huHGFh). This protein is involved in tumour formation in
 CC some cells, and it is thought that the gene and protein will be useful in
 CC the diagnosis and treatment of infections, autoimmune disorders, vascular
 CC diseases and cancers. In addition, the gene can be used to produce
 CC transgenic animals which may be used as animal models for these diseases,
 CC and the protein can be used to induce cell proliferation, haematopoiesis,
 CC lymphocyte proliferation and angiogenesis.

SQ Sequence 203 AA;

Query Match 14.3%; Score 189; DB 21; Length 203;
 Best Local Similarity 28.2%; Pred. No. 2.5e-10;
 Matches 55; Conservative 37; Mismatches 79; Indels 24; Gaps 5;
 QY 7 RLGLDVLWGKLGKRYPPWPGKIVNPPKDLKPRGKCKFFVKGTEGTHAWIKVEQLKPYHAAH 66
 DB 10 KAGDLVFAKMKGYPHWPARIDELPEGAVKPPANK-YPIFFFGTHETAFLGPKOLFPPYKEY 68
 QY 67 KEEMIKINKGRFOQAQDAVE-----EFLRRAGKQDOTSHNSDDKNRNSSEERSRPN 121
 DB 69 KDKFGKSNKRKGFNEGLWEIENNPVGFYQAIQQOSSSETEGEGGNTADASSE----- 123
 QY 122 SGDEKRLKLSISEGKVKKNMGEGKRVSSGSSERGSKSPKRAQEQSPRKRGRPPKDEKOL 181
 DB 124 -----EEGDRVEEDGKGRKNEKAGSKRKSYTSKKSKQSRKSPG--DEDDKDC 171
 QY 182 TTPES-STVKGMMAG 195
 DB 172 KEENKSSSEGGDAG 186

RESULT 15

AA88374
 ID AA88374 standard; Protein; 203 AA.

XX AC AA88374;
 XX AC AA88374;

DT 20-JUL-2000 (first entry)

XX Human type II hepatoma derived growth factor (HDGF2) amino acid sequence.

XX Hepatoma-derived growth factor 2; HDGF2; cytostatic; drug;

KW treatment; hepatoma.

OS Homo sapiens.

XX WO200017351-A1.

PN 30-MAR-2000.

XX 06-SEP-1999; 99WO-CN00139.

XX 22-SEP-1998; 98CN-0119758.

XX (YULL/) YU L.

XX Yu L, Zhang H, Fu Q, Zhao Y, Tu Q;

XX WPI; 2000-283579/24.

DR N-PSDB; AAA13161.

XX Type II human-derived growth factor with homology to its type I
 PT version, useful e.g. in study and development of drugs for hepatoma
 PS Claim 2; Page 15; 27pp; Chinese.

XX This sequence represents the human hepatoma-derived growth factor (HDGF2)
 CC protein sequence. The HDGF2 sequence shows considerable homology to the
 CC type I version. The protein has cytostatic activity. The HDGF2 protein,
 CC derivatives and polynucleotides are useful e.g. in the study and
 CC development of drugs for the treatment of hepatoma.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 8, 2003, 11:09:20 ; Search time 31 Seconds
(without alignments)
1493.869 Million cell updates/sec

Title: US-09-987-755-2

Perfect score: 1319

Sequence: 1 MAASVRLGDLVWGLGRYP.....VCYQAITKKIKICEDLLLR 249

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 684280 seqs, 185983659 residues

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*

2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*

3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*

4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*

5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*

6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*

7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*

8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*

9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*

10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*

11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*

12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*

13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*

14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*

15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*

16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*

17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*

18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1319	100.0	249	11	US-09-987-755-2
2	1319	100.0	260	15	US-10-103-313-474
3	1314	99.6	250	15	US-10-103-313-317
4	1291	97.9	547	12	US-10-067-482-3
5	1291	97.9	553	12	US-10-067-482-2
6	1278	96.9	550	15	US-10-103-313-417
7	1263	95.8	269	15	US-10-103-313-540
8	190.5	14.4	202	11	US-09-866-050A-651
9	189	14.3	203	9	US-09-768-826-41
10	189	14.3	203	12	US-10-247-671-132
11	186.5	14.1	246	9	US-09-925-302-511
12	185.5	14.1	221	12	US-10-418-445-4
13	185.5	14.1	237	9	US-09-938-885A-5
14	185.5	14.1	240	9	US-09-938-885A-3
15	185.5	14.1	240	11	US-09-987-755-7

16	185.5	14.1	240	12	US-10-116-275-220	Sequence 220, App
17	185.5	14.1	240	15	US-10-207-791-2	Sequence 2, Appli
18	178.5	13.5	235	9	US-09-938-885A-1	Sequence 1, Appli
19	165	12.5	314	9	US-09-768-826-43	Sequence 43, Appl
20	165	12.5	670	11	US-09-823-187-86	Sequence 86, Appl
21	165	12.5	670	12	US-09-863-776-51	Sequence 51, Appl
22	165	12.5	671	11	US-09-946-374-308	Sequence 308, App
23	165	12.5	671	11	US-09-823-187-85	Sequence 85, Appl
24	165	12.5	671	12	US-10-015-387A-308	Sequence 308, App
25	165	12.5	671	12	US-10-006-130A-308	Sequence 308, App
26	165	12.5	671	12	US-10-199-672-346	Sequence 346, App
27	165	12.5	671	12	US-10-006-172A-308	Sequence 308, App
28	165	12.5	671	12	US-10-187-749-346	Sequence 346, App
29	165	12.5	671	12	US-10-194-457-346	Sequence 346, App
30	165	12.5	671	12	US-10-184-642-346	Sequence 346, App
31	165	12.5	671	12	US-10-196-747-346	Sequence 346, App
32	165	12.5	671	12	US-10-015-392A-308	Sequence 308, App
33	165	12.5	671	12	US-10-017-253A-308	Sequence 308, App
34	165	12.5	671	12	US-10-173-689-346	Sequence 346, App
35	165	12.5	671	12	US-10-173-690-346	Sequence 346, App
36	165	12.5	671	12	US-10-173-691-346	Sequence 346, App
37	165	12.5	671	12	US-10-173-692-346	Sequence 346, App
38	165	12.5	671	12	US-10-173-694-346	Sequence 346, App
39	165	12.5	671	12	US-10-173-698-346	Sequence 346, App
40	165	12.5	671	12	US-10-173-699-346	Sequence 346, App
41	165	12.5	671	12	US-10-173-707-346	Sequence 346, App
42	165	12.5	671	12	US-10-174-569-346	Sequence 346, App
43	165	12.5	671	12	US-10-174-583-346	Sequence 346, App
44	165	12.5	671	12	US-10-174-587-346	Sequence 346, App
45	165	12.5	671	12	US-10-174-589-346	Sequence 346, App

ALIGNMENTS

RESULT 1

US-09-987-755-2
; Sequence 2, Application US/09987755
; Publication No. US20030022312A1
; GENERAL INFORMATION:
; APPLICANT: Kunsch et al
; TITLE OF INVENTION: Human Hepatoma-Derived Growth Factor-2
; FILE REFERENCE: PF198D1C1
; CURRENT APPLICATION NUMBER: US/09/987,755
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: 09/263,625
; PRIOR FILING DATE: 1999-03-05
; PRIOR APPLICATION NUMBER: 08/464,600
; PRIOR FILING DATE: 1995-06-05
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 249
; TYPE: PRT
; ORGANISM: human
US-09-987-755-2

Query Match	100.0%;	Score 1319;	DB 11;	Length 249;			
Best Local Similarity	100.0%;	Pred. No. 1.3e-102;					
Matches 249;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;			
Qy	1	MAASVRLGDLVWGKLG	RYPPWP	PKIYNPPKDKL	KPRGKKCF	FFVKPFGT	EDHAWIKVEQL 60
Db	1	MAASVRLGDLVWGKLG	RYPPWP	PKIYNPPKDKL	KPRGKKCF	FFVKPFGT	EDHAWIKVEQL 60
Qy	61	KPYHAKHEMIKINKGR	RFQQA	VDAAVEEFL	RRAAKGDKQ	TSSHNSDD	KNRNSSEERSRP 120
Db	61	KPYHAKHEMIKINKGR	RFQQA	VDAAVEEFL	RRAAKGDKQ	TSSHNSDD	KNRNSSEERSRP 120
Qy	121	NSGDEKRLKLSL	SGKVKQNGE	KKRVSSGSS	RRGSKSPL	KRAQESPR	KRGPPPKDEKD 180
Db	121	NSGDEKRLKLSL	SGKVKQNGE	KKRVSSGSS	RRGSKSPL	KRAQESPR	KRGPPPKDEKD 180

Qy	181	LTIPSSSTVKGWAGMGAFAKWKQOPTASFPVKDADPHFHHLLSOTEXPAVCYQAITKKLK	240
Db	181	LTIPSSSTVKGWAGMGAFAKWKQOPTASFPVKDADPHFHHLLSOTEXPAVCYQAITKKLK	240
Qy	241	ICEDLLLP	249
Db	241	ICEDLLLP	249

RESULT 2

```

US-10-103-313-474
; Sequence 474, Application US/10103313
; Publication No. US20030082758A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PZ07C1
; CURRENT APPLICATION NUMBER: US/10/103,313
; CURRENT FILING DATE: 2002-03-12
; NUMBER OF SEQ ID NOS: 653
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 474
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-103-313-474

```

Query Match	100.0%	Score	1319	DB	15	Length	260
Best Local Similarity	100.0%	Pred. No.	1.4e-102				
Matches	249	Conservative	0	Mismatches	0	Indels	0
Qy	1	MAAVSLRLGDLVWGKLG	RYPPWP	PGKIVNPP	KDLKKPRGK	CKFFVK	FCGTEDHAWIKVEQL 60
Db	12	MAAVSLRLGDLVWGKLG	RYPPWP	PGKIVNPP	KDLKKPRGK	CKFFVK	FCGTEDHAWIKVEQL 71
Qy	61	KPYHAHKEEMIK	NGKRFQ	QAVDAV	BEFLBRAK	GKQOTS	SHNSGDDKVRNRSSEERSRP 120
Db	72	KPYHAHKEEMIK	NGKRFQ	QAVDAV	BEFLBRAK	GKQOTS	SHNSGDDKVRNRSSEERSRP 131
Qy	121	NSGDEKRLSL	SEGVKVK	NMGEGKKR	VSSGSS	SERGSK	SPLKRAEQSQSPKRGRPPKDEKD 180
Db	132	NSGDEKRLSL	SEGVKVK	NMGEGKKR	VSSGSS	SERGSK	SPLKRAEQSQSPKRGRPPKDEKD 191
Qy	181	LTPESSTVCKM	AGPMAAFK	WQPTAS	EPVKDADPH	FHHFLL	SOTEKPAVCIQAITKKLK 240
Db	192	LTPESSTVCKM	AGPMAAFK	WQPTAS	EPVKDADPH	FHHFLL	SOTEKPAVCIQAITKKLK 251
Qy	241	ICEDLLLP	R	249			
Db	252	ICEDLLLP	R	260			

DEPT. T. 3

```

RESULI 3
US-10-103-313-317
; Sequence 317, Application US/10103313
; Publication No. US20030082758A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJZ07C1
; CURRENT APPLICATION NUMBER: US/10/103.313
; CURRENT FILING DATE: 2002-03-12
; NUMBER OF SEQ ID NOS: 653
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 317
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-103-313-317

```

Query Match 99.6%; Score 1314; DB 15; Length 250;

Best Local Similarity	100.0%;	Pred. No. 3.4e-102;			
Matches	248;	Conservative	0;	Mismatches	0;
				Indels	0;
				Gaps	0;
Qy	2	AAVSLRLGDLVWGKLGRYPPWPGKI	VNPPPKDLKKPRGKCCFFVKFFCTEDHAMI	KVQLK	61
Db	3	AAVSLRLGDLVWGKLGRYPPWPGKI	VNPPPKDLKKPRGKCCFFVKFFCTEDHAMI	KVQLK	62
Qy	62	PYHAHKEEMIKINKGRFQQA	VDABEEFLRRAKGKQDTSSHNSDDDKNRNRSSEERSRP	N	121
Db	63	PYHAHKEEMIKINKGRFQQA	VDABEEFLRRAKGKQDTSSHNSDDDKNRNRSSEERSRP	N	122
Qy	122	SGDEKRLSLSEGKVKQWGEKKRVSSGSSRGSKS	PLKRAEQSPRKEGRPPKDEKDL		181
Db	123	SGDEKRLSLSEGKVKQWGEKKRVSSGSSRGSKS	PLKRAEQSPRKEGRPPKDEKDL		182
Qy	182	TIPESSTVKGMAAGPMAAFK	WQPTASBPVKDADPHPHHFLLSQTEKPAVCYQAITTKLKI		241
Db	183	TIPESSTVKGMAAGPMAAFK	WQPTASBPVKDADPHPHHFLLSQTEKPAVCYQAITTKLKI		242
Qy	242	CEDLLLP	R	249	
Db	243	CEDLLLP	R	250	

RESULT 4

```

US-10-067-482-3
; Sequence 3, Application US/10067482
; Publication No. US20030148407A1
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies, Inc.
; TITLE OF INVENTION: Angiogenesis Dehydrogenase Gene
; FILE REFERENCE: 1U 102 R1
; CURRENT APPLICATION NUMBER: US/10/067,482
; CURRENT FILING DATE: 2002-02-07
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 547
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-067-482-3

```

Query Match	97.9%	Score 1291;	DB 12;	Length 547;	
Best Local Similarity	99.6%;	Pred. No. 7.5e-100;			
Matches 243;	Conservative 1;	Mismatches 0;	Indels 0;	Gaps 0;	
Qy 1	MAAVSLRLGDLVWGKLGRIYPPWPGKI	VNPKDLKPRGKKCF	VKPF	GGTGDHAWIKVEQL 60	
Db 1	MAAVSLRLGDLVWGKLGRIYPPWPGKI	VNPKDLKPRGKKCF	VKPF	GGTGDHAWIKVEQL 60	
Qy 61	KPHAHKEEMI	KINKGRFOQAVDAVEEF	LURRAKGDDTSSHNSDD	DKNRNSEE	SRP 120
Db 61	KPHAHKEEMI	KINKGRFOQAVDAVEEF	LURRAKGDDTSSHNSDD	DKNRNSEE	SRP 120
Qy 121	NSGDEKRLSL	SEGKVKVKNNGEGKRVSSGSS	BERGSKSPLKRAEQ	SPKRGPPPKDEK	180
Db 121	NSGDEKRLSL	SEGKVKVKNNGEGKRVSSGSS	BERGSKSPLKRAEQ	SPKRGPPPKDEK	180
Qy 181	LTIPESSTVKGMMAGPMAAFKWOPT	ASEPVKDADPHFHFL	LSQTEKPAVCYQA	ITTKLK 240	
Db 181	LTIPESSTVKGMMAGPMAAFKWOPT	ASEPVKDADPHFHFL	LSQTEKPAVCYQA	ITTKLK 240	
Qy 241	ICED 244				
Db 241	ICBE 244				

RESIST

RESUL 3
US-10-067-482-2
; Sequence 2, Application US/10067482
; Publication No. US20030148407A1
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies, Inc.

; TITLE OF INVENTION: Angiogenesis Dehydrogenase Gene
; FILE REFERENCE: 1U 102 R1
; CURRENT APPLICATION NUMBER: US/10/067,482
; CURRENT FILING DATE: 2002-02-07
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 553
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-067-482-2

Query Match 97.9%; Score 1291; DB 12; Length 553;
Best Local Similarity 99.6%; Pred. No. 7.6e-100;
Matches 243; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MAASLRLGLVWGLGRYPWPGKIYNPPKDLKKPRGKCFVVKFGTGDHAWIKVEQL 60
Db 1 MAASLRLGLVWGLGRYPWPGKIYNPPKDLKKPRGKCFVVKFGTGDHAWIKVEQL 60
Qy 61 KPYHAHEEMIKKNGKRFQQAQVDAVEEFLRRRAKGDQDTSSHNSDDKNRRNSEERSRP 120
Db 61 KPYHAHEEMIKKNGKRFQQAQVDAVEEFLRRRAKGDQDTSSHNSDDKNRRNSEERSRP 120
Qy 121 NSGDEKELSLSEGKVKNNMGEGKRVSSGSSERGSKSPKRAQEQSPKRGPRPPKDEKD 180
Db 121 NSGDEKELSLSEGKVKNNMGEGKRVSSGSSERGSKSPKRAQEQSPKRGPRPPKDEKD 180
Qy 181 LTIPESVTGMMAGPMAAFKQWPTASEPVKADDPHFFHLLSQTEKPAVCYQAITKKLK 240
Db 181 LTIPESVTGMMAGPMAAFKQWPTASEPVKADDPHFFHLLSQTEKPAVCYQAITKKLK 240
Qy 241 ICED 244
Db 241 ICED 244

RESULT 6
US-10-103-313-417
; Sequence 417, Application US/10103313
; Publication No. US20030082758A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P207C1
; CURRENT APPLICATION NUMBER: US/10/103,313
; CURRENT FILING DATE: 2002-03-12
; NUMBER OF SEQ ID NOS: 653
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 417
; LENGTH: 550
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-103-313-417

Query Match 96.9%; Score 1278; DB 15; Length 550;
Best Local Similarity 99.6%; Pred. No. 9.2e-99;
Matches 240; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 4 VSLRLGLVWGLGRYPWPGKIYNPPKDLKKPRGKCFVVKFGTGDHAWIKVEQLKPY 63
Db 7 VSLRLGLVWGLGRYPWPGKIYNPPKDLKKPRGKCFVVKFGTGDHAWIKVEQLKPY 66
Qy 64 HAHKEEMIKKNGKRFQQAQVDAVEEFLRRRAKGDQDTSSHNSDDKNRRNSEERSRPNSG 123
Db 67 HAHKEEMIKKNGKRFQQAQVDAVEEFLRRRAKGDQDTSSHNSDDKNRRNSEERSRPNSG 126
Qy 124 DEKRLSLSEGKVKNNMGEGKRVSSGSSERGSKSPKRAQEQSPKRGPRPPKDEKDLTI 183
Db 127 DEKRLSLSEGKVKNNMGEGKRVSSGSSERGSKSPKRAQEQSPKRGPRPPKDEKDLTI 186
Qy 184 PESSTVKGMMAGPMAAFKQWPTASEPVKADDPHFFHLLSQTEKPAVCYQAITKKLKICE 243

Db 187 PESSTVKGMMAGPMAAFKQWPTASEPVKADDPHFFHLLSQTEKPAVCYQAITKKLKICE 246
Qy 244 D 244
Db 247 E 247
RESULT 7
US-10-103-313-540
; Sequence 540, Application US/10103313
; Publication No. US20030082758A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P207C1
; CURRENT APPLICATION NUMBER: US/10/103,313
; CURRENT FILING DATE: 2002-03-12
; NUMBER OF SEQ ID NOS: 653
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 540
; LENGTH: 269
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: misc feature
; LOCATION: (165)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-103-313-540

Query Match 95.8%; Score 1263; DB 15; Length 269;
Best Local Similarity 98.8%; Pred. No. 6.9e-98;
Matches 238; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 4 VSLRLGLVWGLGRYPWPGKIYNPPKDLKKPRGKCFVVKFGTGDHAWIKVEQLKPY 63
Db 1 VSLRLGLVWGLGRYPWPGKIYNPPKDLKKPRGKCFVVKFGTGDHAWIKVEQLKPY 60
Qy 64 HAHKEEMIKKNGKRFQQAQVDAVEEFLRRRAKGDQDTSSHNSDDKNRRNSEERSRPNSG 123
Db 61 HAHKEEMIKKNGKRFQQAQVDAVEEFLRRRAKGDQDTSSHNSDDKNRRNSEERSRPNSG 120
Qy 124 DEKRLSLSEGKVKNNMGEGKRVSSGSSERGSKSPKRAQEQSPKRGPRPPKDEKDLTI 183
Db 121 DEKRLSLSEGKVKNNMGEGKRVSSGSSERGSKSPKRAQEQSPKRGPRPPKDEKDLTI 180
Qy 184 PESSTVKGMMAGPMAAFKQWPTASEPVKADDPHFFHLLSQTEKPAVCYQAITKKLKICE 243
Db 181 PESSTVKGMMAGPMAAFKQWPTASEPVKADDPHFFHLLSQTEKPAVCYQAITKKLKICE 240
Qy 244 D 244
Db 241 E 241

RESULT 8
US-09-866-050A-651
; Sequence 651, Application US/09866050A
; Publication No. US2003004071A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James G.
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods for Their Use
; FILE REFERENCE: 11000.1011c4U

Query Match 14.3%; Score 189; DB 12; Length 203;
Best Local Similarity 28.2%; Pred. No. 3.8e-08;

Qy	7	RLGDLVWGKLGRIYIPWPGKIVNPPKDLKPKPGKKCFVKKFFGTEDHAIKIVKQLKPYHAAH	66
Db	10	KAGDLVFAKMKGYPHWPARIDELPEGAVKPPANK-YPIFFGTHTAFLGPKDLPPYKEY	68
Qy	67	KEEMIKNKNGKRFQQAQDAVE----BFLRRAGKDOTSNNSSDDKNRRNSSERSRPN	121
Db	69	KDFGKNKNGKGFNEGWEIENNPGVFTGYQALQQSSSTEGEGGNTADASSE-----	123
Qy	122	SGDEKRLKLSLEGVKKNMGEGKKRVSSGSSGSKSPFLKRAQBSQPRKGRPPKDEKDL	181
Db	124	-----EEGRDVEEDCKGKKNKAGSKRKKS YTSKKSQKRSKPG--DEDDKDC	171
Qy	182	TIPES-STVKGMMAG	195
Db	172	KEENKSSSEGGDAG	186

```

RESULT 11
US-09-925-302-511
; Sequence 511, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 511
; LENGTH: 246
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: SITE
; LOCATION: (207)
; OTHER INFORMATION: Xaa equals any of the naturally occurring
US-09-925-302-511

```

Query Match 14.1%; Score 186.5; DB 9; Length 246;
Best Local Similarity 29.8%; Pred. No. 7.8e-08;
Matches 64; Conservative 30; Mismatches 100; Indels 21; Gaps 8;

QY 7 RLGLVWGLGRYPWPVKIWNPPKDLKXPRGKCFVKEFGTETHAWIKVQOLKPYH 66
DB 17 KCGDLVFAKMGYPHWPARDIEMPEAAVKSTANK-YQVFFFGTHETAFGLPKOLF 75
QY 67 KEEMIKNGKRFQQAQVDAVEEFLRRAKGDKQTSNHNSSDDKNRNSSEERSRPN 126
DB 76 KEKFGPNKRGKGFSEGLWEIEN-----NPTVKASGYQSSQKSCVPEPEPEAE 130
QY 127 RKLSEGVKKNMGEGKRVSSGSSRGSKPLKRAQ-----EQSPR--KRGPPKDE 180
DB 131 KK-GNAEGSSDE---EGKLVIDEPAKEKNEKALKRRAGDLLEDSPKPKAEAP 185
QY 181 LTIPESSTVKGMMAGPMAAFKQPTASEPVKADDP 215
DB 186 ---KEATLEVERPLPMEVEK-NSTXSEPGSGRGP 216

RESULT 12
US-10-418-445-4
; Sequence 4, Application US/10418445
; Publication No. US20030190708A1
; GENERAL INFORMATION:
; APPLICANT: BURGESS, et al.
; TITLE OF INVENTION: No. US20030190708A1el Hepatoma-Derived Growth Factor-Like Protein
; FILE REFERENCE: Polynucleotides Encoding Them and Methods of Use
; CURRENT APPLICATION NUMBER: US/10/418,445
; PRIOR FILING DATE: 2003-04-17
; PRIOR FILING DATE: 2000-09-28
; PRIOR FILING DATE: 2000-09-28
; PRIOR FILING DATE: 1999-10-01
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 221
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-418-445-4

Query Match 14.1%; Score 185.5; DB 12; Length 221;
Best Local Similarity 29.8%; Pred. No. 8.3e-08;
Matches 64; Conservative 30; Mismatches 100; Indels 21; Gaps 8;

QY 7 RLGLVWGLGRYPWPVKIWNPPKDLKXPRGKCFVKEFGTETHAWIKVQOLKPYH 66
DB 2 KCGDLVFAKMGYPHWPARDIEMPEAAVKSTANK-YQVFFFGTHETAFGLPKOLF 60
QY 67 KEEMIKNGKRFQQAQVDAVEEFLRRAKGDKQTSNHNSSDDKNRNSSEERSRPN 126
DB 61 KEKFGPNKRGKGFSEGLWEIEN-----NPTVKASGYQSSQKSCVPEPEPEAE 115
QY 127 RKLSEGVKKNMGEGKRVSSGSSRGSKPLKRAQ-----EQSPR--KRGPPKDE 180
DB 116 KK-GNAEGSSDE---EGKLVIDEPAKEKNEKALKRRAGDLLEDSPKPKAEAP 170
QY 181 LTIPESSTVKGMMAGPMAAFKQPTASEPVKADDP 215
DB 171 ---KEATLEVERPLPMEVEK-NSTXSEPGSGRGP 201

RESULT 13
US-09-938-885A-5
; Sequence 5, Application US/09938885A
; Patent No. US20020090679A1
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; Goli, Surya K.

Query Match 14.1%; Score 185.5; DB 9; Length 237;
Best Local Similarity 29.8%; Pred. No. 9.1e-08;
Matches 65; Conservative 32; Mismatches 94; Indels 27; Gaps 9;

QY 7 RLGLVWGLGRYPWPVKIWNPPKDLKXPRGKCFVKEFGTETHAWIKVQOLKPYH 66
DB 11 KCGDLVFAKMGYPHWPARDIEMPEAAVKSTANK-YQVFFFGTHETAFGLPKOLF 69
QY 67 KEEMIKNGKRFQQAQVDAVEEFLRRAKGDKQTSNHNSSDDKNRNSSEERSRPN 126
DB 70 KEKFGPNKRGKGFSEGLWEIEN-----NPTVKASGYQSSQKSCVPEPEAE 124
QY 127 RKLSEGVKKNMGEGKRVSSGSSRGSKPLKRAQ-----EQSPR--KRGPPK 177
DB 125 KKGS-AEGSSDE---EGKLVIDEPAKEKNEKALKRRAGDLLEDSPKPK 176
QY 178 EKDLTPESSTVKGMMAGPMAAFKQPTASEPVKADDP 215
DB 177 EED---KEATLEVERPLPMEVEK-NSTXSEPGSGRGP 210

RESULT 14
US-09-938-885A-3
; Sequence 3, Application US/09938885A
; Patent No. US20020090679A1
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; Goli, Surya K.
; Murry, Lynn E.

TITLE OF INVENTION: LUNG GROWTH FACTOR VARIANT
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/938,885A
FILING DATE: 24-Aug-2001
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/760,745
FILING DATE: 1996-12-07
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0169 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 237 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 945419
SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-938-885A-5

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 8, 2003, 11:07:05 ; Search time 35 Seconds
(without alignments)
1835.860 Million cell updates/sec

Title: US-09-987-755-2
Perfect score: 1319
Sequence: 1 MAASVRLGLVWGKGRYP.....VCYQAITKKIKICEDLLPR 249

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL 23.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_virus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1291	97.9	547	Q8BXX2	Q8bxx2 homo sapien
2	1247.5	94.6	546	Q922P9	Q922p9 mus musculu
3	1225.5	92.9	546	Q9CYQ1	Q9cyq1 mus musculu
4	1138	86.3	523	Q9BT11	Q9bt11 homo sapien
5	201.5	15.3	239	Q9XSK7	Q9xsk7 bos taurus
6	193	14.6	475	Q9VAA9	Q9vaa9 drosophila
7	191.5	14.5	202	Q9JMG7	Q9jmg7 mus musculu
8	190.5	14.4	202	Q923W4	Q923w4 rattus norv
9	190.5	14.4	202	Q9D2M7	Q9d2m7 mus musculu
10	190.5	14.4	205	Q8BQ69	Q8bq69 mus musculu
11	189	14.3	203	Q9Y3E1	Q9y3e1 homo sapien
12	186.5	14.1	237	Q8BPG7	Q8bpg7 mus musculu
13	186.5	14.1	603	Q8T079	Q8t079 drosophila
14	186.5	14.1	603	Q9VLS1	Q9vls1 drosophila
15	184.5	14.0	237	Q8VHK7	Q8vhk7 rattus norv
16	184.5	14.0	237	Q923W3	Q923w3 rattus norv

ALIGNMENTS

RESULT 1

Q9BXX2 ID Q9BXX2 PRELIMINARY; PRT; 547 AA.

AC Q9BXX2; 01-JUN-2001 (TrEMBLrel. 17, Created)

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Cytokine-like nuclear factor n-pac.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA New L., Han J.;

RT "A novel cytokine-like nuclear factor, n-pac.";

RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF326966; AAK15524.1; -

DR InterPro; IPR006115; 6FGD_NAD.

DR InterPro; IPR000637; AT_hook.

DR InterPro; IPR000313; PWWP domain.

DR Pfam; PF02178; AT_hook; 1.

DR Pfam; PF01446; NAD binding_2; 1.

DR Pfam; PF00855; PWWP; 1.

DR SMART; SM00384; AT_hook; 1.

DR PROSITE; PS00812; PWWP; 1.

SQ SEQUENCE 547 AA; 59827 MW; C7D785CCBF83204A CRC64;

Query Match 97.9%; Score 1291; DB 4; Length 547;

Best Local Similarity 99.6%; Pred. No. 1.5e-97;

Matches 243; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAASVRLGLVWGKGRYPFPFKI VPPKDLKKPRGKCFVFKFPGTGDHAWIKVEQL 60

Db 1 MAASVRLGLVWGKGRYPFPFKI VPPKDLKKPRGKCFVFKFPGTGDHAWIKVEQL 60

Qy 61 KPVHAHKEIMKINKGRFQQA VDAVEEFRRAGKQDTSHNSDDKXNRNSESRRP 120

Db 61 KPVHAHKEIMKINKGRFQQA VDAVEEFRRAGKQDTSHNSDDKXNRNSESRRP 120

```
QY 121 NSGDEKRLSLSEGKVKKNMGKKRVSSGSSGSKSPKRAQSQSPKRGPRPKDEK 180
DB 121 NSGDEKRLSLSEGKVKKNMGKKRVSSGSSGSKSPKRAQSQSPKRGPRPKDEK 180
QY 181 LTPESSTVKGMMAGPMAAFKQWPTASEPVKDADPHFHFLLSQTEKPAVCYQAITTKLK 240
DB 181 LTPESSTVKGMMAGPMAAFKQWPTASEPVKDADPHFHFLLSQTEKPAVCYQAITTKLK 240
QY 241 ICED 244
DB 241 ICEE 243

RESULT 2
Q922P9
ID Q922P9 PRELIMINARY; PRT; 546 AA.
AC Q922P9;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE RIKEN GDNA 393040IK13RIK.
GN 393040IK13RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC006893; AAH06893.1; -.
DR MGI; MGI:1921272; 393040IK13RIK.
DR InterPro; IPR006115; 6PGD_NAD.
DR InterPro; IPR00637; AT hook.
DR InterPro; IPR00313; PWWP domain.
DR Pfam; PF02178; AT hook; 1.
DR Pfam; PF03446; NAD binding_2; 1.
DR Pfam; PF00855; PWWP; 1.
DR SMART; SM00384; AT_hook; 1.
DR PROSITE; PS0812; PWWP; 1.
SQ SEQUENCE 546 AA; 59715 MW; F5D2090DE1F64723 CRC64;

Query Match 94.6%; Score 1247.5; DB 11; Length 546;
Best Local Similarity 96.3%; Pred. No. 5.5e-94;
Matches 235; Conservative 6; Mismatches 2; Indels 1; Gaps 1;

QY 1 MAASLRLGDLVWGKLGKRYPPWPGKIVNPPKDLKKPRGKCKFFVKFFGTEDHAWIKVEQL 60
DB 1 MAASLRLGDLVWGKLGKRYPPWPGKIVNPPKDLKKPRGKCKFFVKFFGTEDHAWIKVEQL 60
QY 61 KPYHAHKEEMIKKNGKRFQQAQVDAVEEFLRAKGDQTSNHSDDKNNRNSSEERSRP 120
DB 61 KPYHAHKEEMIKKNGKRFQQAQVDAVEEFLRAKGDQTSNHSDDKNNRNSSEERSRP 120
QY 121 NSGDEKRLSLSEGKVKKNMGKKRVSSGSSGSKSPKRAQSQSPKRGPRPKDEK 180
DB 121 NSGDEKRLSLSEGKVKKNMGKKRVSSGSSGSKSPKRAQSQSPKRGPRPKDEK 179
QY 181 LTPESSTVKGMMAGPMAAFKQWPTASEPVKDADPHFHFLLSQTEKPAVCYQAITTKLK 240
DB 181 LTPESSTVKGMMAGPMAAFKQWPTASEPVKDADPHFHFLLSQTEKPAVCYQAITTKLK 239
QY 241 ICED 244
DB 240 ICEE 243

RESULT 3
Q9CYQ1
ID Q9CYQ1 PRELIMINARY; PRT; 546 AA.
AC Q9CYQ1;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
```

```
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
GN 393040IK13RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC006893; AAH06893.1; -.
DR MGI; MGI:1921272; 393040IK13RIK.
DR InterPro; IPR006115; 6PGD_NAD.
DR InterPro; IPR00637; AT hook.
DR InterPro; IPR00313; PWWP domain.
DR Pfam; PF02178; AT hook; 1.
DR Pfam; PF03446; NAD binding_2; 1.
DR Pfam; PF00855; PWWP; 1.
DR SMART; SM00384; AT_hook; 1.
DR PROSITE; PS0812; PWWP; 1.
SQ SEQUENCE 546 AA; 59744 MW; B3AC1562477ABC1F CRC64;

Query Match 92.98%; Score 1225.5; DB 11; Length 546;
Best Local Similarity 94.7%; Pred. No. 3.5e-92;
Matches 231; Conservative 9; Mismatches 3; Indels 1; Gaps 1;

QY 1 MAASLRLGDLVWGKLGKRYPPWPGKIVNPPKDLKKPRGKCKFFVKFFGTEDHAWIKVEQL 60
DB 1 MAASLRLGDLVWGKLGKRYPPWPGKIVNPPKDLKKPRGKCKFFVKFFGTEDHAWIKVEQL 60
QY 61 KPYHAHKEEMIKKNGKRFQQAQVDAVEEFLRAKGDQTSNHSDDKNNRNSSEERSRP 120
DB 61 KPYHAHKEEMIKKNGKRFQQAQVDAVEEFLRAKGDQTSNHSDDKNNRNSSEERSRP 120
QY 121 NSGDEKRLSLSEGKVKKNMGKKRVSSGSSGSKSPKRAQSQSPKRGPRPKDEK 180
DB 121 NSGDEKRLSLSEGKVKKNMGKKRVSSGSSGSKSPKRAQSQSPKRGPRPKDEK 179
QY 181 LTPESSTVKGMMAGPMAAFKQWPTASEPVKDADPHFHFLLSQTEKPAVCYQAITTKLK 240
DB 181 LTPESSTVKGMMAGPMAAFKQWPTASEPVKDADPHFHFLLSQTEKPAVCYQAITTKLK 239
QY 241 ICED 244
DB 240 ICEE 243

RESULT 4
Q9BT11
ID Q9BT11 PRELIMINARY; PRT; 523 AA.
AC Q9BT11;
```

DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Similar to RIKEN cDNA 3930401K13 gene (Fragment).
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RA Strausberg R.;
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC003693; AAH03693.1;
 DR InterPro; IPR006115; GPGD_NAD.
 DR InterPro; IPR000637; AT hook.
 DR InterPro; IPR000313; PWWP_domain.
 DR Pfam; PF02178; AT hook; 1.
 DR Pfam; PF03446; NAD binding_2; 1.
 DR Pfam; PF00855; PWWP; 1.
 DR SMART; SM00384; AT hook; 1.
 DR PROSITE; PS50812; PWWP; 1.
 FT NON_TER 1
 SQ SEQUENCE 523 AA; 57215 MW; ED3E2853938A211F CRC64;
 Query Match 86.3%; Score 1138; DB 4; Length 523;
 Best Local Similarity 100.0%; Pred. No. 4.8e-85;
 Matches 214; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 14 GKLGRYPWPVKIIVNPPKDLKPRGKKCFVKGFGTGDHAWIKVQLKPYHAKHEMIKI 73
 DB 1 GKLGRYPWPVKIIVNPPKDLKPRGKKCFVKGFGTGDHAWIKVQLKPYHAKHEMIKI 60
 QY 74 NKCKRQQQAVDAVEEFLRRAGKQDQTSNHNSSDDKNRRNSSERSRPNSSGDEKRLSLSE 133
 DB 61 NKCKRFQQAQVDAVEEFLRRAGKQDQTSNHNSSDDKNRRNSSERSRPNSSGDEKRLSLSE 120
 QY 134 GYVKKNMGEKGRVSSGSSSRGSKSPLKRAQEOSPRKGRGPPPKDKDLTTPESSTVKGMM 193
 DB 121 GYVKKNMGEKGRVSSGSSSRGSKSPLKRAQEOSPRKGRGPPPKDKDLTTPESSTVKGMM 180
 QY 194 AGPMAAFKWPQTAPEPVKDADPHFHLFSQTEK 227
 DB 181 AGPMAAFKWPQTAPEPVKDADPHFHLFSQTEK 214
 RESULT 5
 ID Q9XSK7 PRELIMINARY; PRT; 239 AA.
 AC Q9XSK7;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Hepatoma derived growth factor.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RA Dietz F., Nakamura H., Gieselmann V.;
 RL "HRP-3: A new member of the hepatoma derived growth factor related protein family interacts with HDGF and another HDGF related polypeptide."
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ237996; CAB40626.1; --
 DR InterPro; IPR000313; PWWP_domain.
 DR Pfam; PF00855; PWWP; 1.
 DR SMART; SM00293; PWWP; 1.
 DR PROSITE; PS50812; PWWP; 1.

SQ SEQUENCE 239 AA; 26604 MW; 6EA42D4D2FBDAP9E CRC64;
 Query Match 15.3%; Score 201.5; DB 6; Length 239;
 Best Local Similarity 30.7%; Pred. No. 8.7e-09;
 Matches 66; Conservative 31; Mismatches 97; Indels 21; Gaps 8;
 QY 7 RLGDVLVWGKLGKRYPPWPKIIVNPPKDLKPRGKKCFVKGFGTGDHAWIKVQLKPYHAKHEMIKI 66
 DB 11 KCGDLVFAKMGKGYPHWPARIDEMPEAAVKSTANK-YQVFFFGTGTAFGLPKDLFFVEES 69
 QY 67 KEEMIKNKGRFQQQAVDAVEEFLRRAGKQDQTSNHNSSDDKNRRNSSERSRPNSSGDEK 126
 DB 70 KEKFGKPNRRKGFSEGLWEIEN-----NPTVKASGYQSSQKSCVEPEPEPEATSGDGD 124
 QY 127 RKLSLSEGVKKNMGEKGRVSSGSSSRGSKSPLKRAQ-----EQSPRKRGRPPKDKDL- 181
 DB 125 KK-GNAGSSDE---EGKLVIDEPTKEKNEKALKRAGDLDLSDSPK-----PKEADLE 176
 QY 182 -TIPESSTVKGMMAGPMAAFKWPQTAPEPVKDADP 215
 DB 177 GEEKATLEGERPLPVEAEK-NSTPEPGSGRGP 210
 RESULT 6
 ID Q9VAA9 PRELIMINARY; PRT; 475 AA.
 AC Q9VAA9;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE CG7946 protein (LD23804P).
 GN CG7946.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,

```
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RC Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RC Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003772; AAF57004.1; -
DR EMBL; AY051690; AAK93114.1; -
DR FlyBase; FBgn0039743; CG7946.
DR InterPro; IPR000313; PWM domain.
DR Pfam; PF00855; PWM; 1.
DR PROSITE; PS50812; PWM; 1.
SQ SEQUENCE 475 AA; 52838 MW; 4A0186A4121F05DC CRC64;
Query Match 14.6%; Score 193; DB 5; Length 475;
Best Local Similarity 27.7%; Pred. No. 9.7e-08;
Matches 64; Conservative 32; Mismatches 89; Indels 46; Gaps 8;
QY 2 AAVSLRLGLVWGKLGRRYPPGKIVNPPKDKLKKPRGKCFVFKFGTETHAWIKVEQLK 61
DB 6 AAASYIGDLVFAKVGYPGPPAKT-----KSNNNKKNVYFGTGTANIKLEDLF 58
QY 62 PVHAHEEMI--KINKGRFQOAVDAVEFLR-----RAKGDQDTSSHNSDDKNRR 111
DB 59 PVASKEFATEKINWRAKFAIEAIOIESALRGESAPIDLLDGAEPVAPPTGD----- 113
QY 112 NSSEERSPNSGDEKRLSLSEGVKQNMGEKGRVSS-----GSSRSGSKSPUKRAQE 165
DB 114 GVKTEPKPEPGPEPATAAPVAAAEEKPKSGTRTKAPPRHVDGDSAGAE----- 166
QY 166 QSPKRGRRPKDEKDLTTPESVTKGMAGPMAAFKWOPTAS-EPVKDADP 215
DB 167 -----APPKARRRVPTTEGLATAAAIAPAAPAA---TPTSSKSKSVKSKP 207
RESULT 7
QJUNG7 PRELIMINARY; PRT; 202 AA.
AC QJUNG7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hepatoma-derived growth factor-related protein HRP-3.
GN HDGFRP3 OR HRP-3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC Ikegama K., Nakamura H.;
RC TISSUE=Testis;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=20050055; PubMed=10581169;
RA Ikegama K., Yamamoto M., Kishima Y., Enomoto H., Yoshida K.,
RA Suemura M., Kishimoto T., Nakamura H.;
RT "A new member of a hepatoma-derived growth factor gene family can
RT translocate to the nucleus.";
RL Biochem. Biophys. Res. Commun. 266:81-87 (1999).
DR EMBL; AB029493; BAA30478.1; -
DR MGD; MGI:1352760; Hdgfrp3.
```

```
DR InterPro; IPR000313; PWM domain.
DR Pfam; PF00855; PWM; 1.
DR SMART; SM00293; PWM; 1.
DR PROSITE; PS50812; PWM; 1.
SQ SEQUENCE 202 AA; 22490 MW; FCD31A0F79AB8D6 CRC64;
Query Match 14.5%; Score 191.5; DB 11; Length 202;
Best Local Similarity 29.1%; Pred. No. 4.7e-08;
Matches 58; Conservative 32; Mismatches 88; Indels 21; Gaps 5;
QY 7 RLGLVWGKLGRRYPPGKIVNPPKDKLKKPRGKCFVFKFGTETHAWIKVEQLKPYHAH 66
DB 10 KAGDLVFAKMGYPHWPARIDELPGAVKPPANK-YPIFFFGTETHAFGLGPKDLFPYKEY 68
QY 67 KEEMIKINKGRFQOAVDAVE-----EFLRAKGDQDTSSHNSDDKNRRNS--EERSR 119
DB 69 KDFGKSNKRGFNEGLWEIENPGVKTGTQTTQQSSSETEGEGGNTADASSEEGDR 128
QY 120 PMSGDEKRLSLSEGVKQNMGEKGRVSSGSSRSGSKSP-----LKRAEQSPRKRG 172
DB 129 VEDGKGRKNEKGGSKKSY-----TSKSSKQSRKSPGDEDDKCKEENKSSSEGG 182
QY 173 RPPKDEKDLTTPESSTVKG 191
DB 183 DAGNDRTRNTTADLOKTSEG 201
RESULT 8
QJ23W4 PRELIMINARY; PRT; 202 AA.
AC QJ23W4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE HRP3.
GN HRP3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Yu L.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF389347; AAK72965.1; -
DR InterPro; IPR000313; PWM domain.
DR Pfam; PF00855; PWM; 1.
DR SMART; SM00293; PWM; 1.
DR PROSITE; PS50812; PWM; 1.
SQ SEQUENCE 202 AA; 22446 MW; FCCDEBBD2BEAB8D6 CRC64;
Query Match 14.4%; Score 190.5; DB 11; Length 202;
Best Local Similarity 28.7%; Pred. No. 5.7e-08;
Matches 56; Conservative 37; Mismatches 77; Indels 25; Gaps 6;
QY 7 RLGLVWGKLGRRYPPGKIVNPPKDKLKKPRGKCFVFKFGTETHAWIKVEQLKPYHAH 66
DB 10 KAGDLVFAKMGYPHWPARIDELPGAVKPPANK-YPIFFFGTETHAFGLGPKDLFPYKEY 68
QY 67 KEEMIKINKGRFQOAVDAVE-----EFLRAKGDQDTSSHNSDDKNRRNSSEERSRPN 121
DB 69 KDFGKSNKRGFNEGLWEIENPGVKTGTQTTQQSSSETEGEGGNTADASSE-----E 124
QY 122 SGDEKRLSLSEGVKQNMGEKGRVSSGSSRSGSKSPFLKRAEQSPRKGRPKDEKDL 181
DB 125 EGD-----RVEDGKGRKNEKGGSKKSYTSKSKQSRKSPG--DEDDKOC 170
QY 182 TTPES-STVKGMMAG 195
DB 171 KEENKSSSEGGDAG 185
RESULT 9
```

```
Q9D2M7
AC Q9D2M7; PRELIMINARY; PRT; 202 AA.
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hepatoma-derived growth factor, related protein 3.
GN HDGFRP3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Skin;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J.,
RA Schirni L.M., Staebli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RN "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK019487; BAB31754.1; -
DR EMBL; AK044858; BAC32123.1; -
DR MGD; MGI:1352760; Hdgrp3.
DR InterPro; IPR00313; PWWP_domain.
DR Pfam; PF00855; PWWP; 1.
DR SMART; SM00293; PWWP; 1.
DR PROSITE; PS00812; PWWP; 1.
SQ SEQUENCE 202 AA; 22430 MW; FCD31BD2BEAB8D6 CRC64;

Query Match 14.4%; Score 190.5; DB 11; Length 202;
Best Local Similarity 28.7%; Pred. No. 5.7e-08;
Matches 56; Conservative 37; Mismatches 77; Indels 25; Gaps 6;

Qy 7 RLGLVWGKLGYPVPPGKIVNPKDKLPKRGKCFVFKFGTGDHAWIKVQOLKPYHAA 66
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 10 KAGDLVFAKMGYPHPWPARIDELPEGAVKPPANK-YPIFFFGTHTAFLGPKDLFPYKEY 68
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Qy 67 KEEMIKNGKRFOQAVDAVE-----EFLRAKGDQTSNHSDDKRRNSSEERSRPN 121
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 69 KDFGKSNKRKGFNEGLWEIENNPVGKFTGYQTIOQSSSETEGEGNTADASSE-----E 124
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Qy 122 SGDEKRLKLSLGKVKKNMGEGKRVSSGSGSERGSKPLKRAQBSKPRKGRPPKDKDL 181
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 125 EGD-----RVEDGKGRKNEKGGSKRKYTSKSKQSRKSPG--DEDDKDC 170
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Qy 182 TTPES-STVKGMMAG 195
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 171 KEENKSSSGGDAG 185
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 11
Q9Y3E1
ID Q9Y3E1 PRELIMINARY; PRT; 203 AA.
AC Q9Y3E1;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE CGI-142 protein (Hypothetical protein FLJ10418) (HRP-3) (Hepatoma-
DE derived growth factor 2).
GN HRP-3 OR HDGF2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Lin W.-C.;
RT "Comparative gene cloning: Identification of novel human genes with
RT Caenorhabditis elegans proteome as template.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Isegai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
```

```
RESULT 10
Q8BQ69
ID Q8BQ69 PRELIMINARY; PRT; 205 AA.
AC Q8BQ69;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hepatoma-derived growth factor.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Dorsal root ganglion;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK051401; BAC34628.1; -
DR EMBL; AK051401; BAC34628.1; -
SQ SEQUENCE 205 AA; 22819 MW; E2AAE3FBB31BD2B CRC64;

Query Match 14.4%; Score 190.5; DB 11; Length 205;
Best Local Similarity 28.7%; Pred. No. 5.8e-08;
Matches 56; Conservative 37; Mismatches 77; Indels 25; Gaps 6;

Qy 7 RLGLVWGKLGYPVPPGKIVNPKDKLPKRGKCFVFKFGTGDHAWIKVQOLKPYHAA 66
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 10 KAGDLVFAKMGYPHPWPARIDELPEGAVKPPANK-YPIFFFGTHTAFLGPKDLFPYKEY 68
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Qy 67 KEEMIKNGKRFOQAVDAVE-----EFLRAKGDQTSNHSDDKRRNSSEERSRPN 121
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 69 KDFGKSNKRKGFNEGLWEIENNPVGKFTGYQTIOQSSSETEGEGNTADASSE-----E 124
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Qy 122 SGDEKRLKLSLGKVKKNMGEGKRVSSGSGSERGSKPLKRAQBSKPRKGRPPKDKDL 181
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 125 EGD-----RVEDGKGRKNEKGGSKRKYTSKSKQSRKSPG--DEDDKDC 170
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Qy 182 TTPES-STVKGMMAG 195
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 171 KEENKSSSGGDAG 185
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 11
Q9Y3E1
ID Q9Y3E1 PRELIMINARY; PRT; 203 AA.
AC Q9Y3E1;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE CGI-142 protein (Hypothetical protein FLJ10418) (HRP-3) (Hepatoma-
DE derived growth factor 2).
GN HRP-3 OR HDGF2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Lin W.-C.;
RT "Comparative gene cloning: Identification of novel human genes with
RT Caenorhabditis elegans proteome as template.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Isegai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
```

RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.,
RT "NEDO human cDNA sequencing project.",
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=20050055; PubMed=10581169;
RA Ikegami K., Yamamoto M., Kishima Y., Enomoto H., Yoshida K.,
RA Suetama M., Kishimoto T., Nakamura H.,
RT "A new member of a hepatoma-derived growth factor gene family can
RT translocate to the nucleus.",
RL Biochem. Biophys. Res. Commun. 266:81-87(1999).
RN [4]
RP SEQUENCE FROM N.A.
RC Tissue=Uterus;
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA Yu L., Fu Q., Tu Q.;
RT "Cloning of a novel human cDNA which is a homolog to mouse hepatoma-
RT derived growth factor (mHDF) and termed hHDF2.",
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF151900; AAD34137.1; -;
DR EMBL; AK001280; BAA91597.1; -;
DR EMBL; AB029156; BAA90477.1; -;
DR EMBL; BC015483; AAH15483.1; -;
DR EMBL; AF110642; AAM27001.1; -;
DR InterPro; IPR000313; PWWP_domain.
DR Pfam; PF00855; PWWP; 1.
DR PROSITE; PS0812; PWWP; 1.
KW Hypothetical protein.
SQ SEQUENCE 203 AA; 22620 MW; 0B660D665F01659C CRC64;

Query Match 14.3%; Score 189; DB 4; Length 203;
Best Local Similarity 28.2%; Pred. No. 7.6e-08;
Matches 55; Conservative 37; Mismatches 79; Indels 24; Gaps 5;

QY 7 RLGLDVLWGKLGKRYPPWPKGVNPPKDLKKPRGKCFVKKFFGTEDHAWIKVEQLKPYHAH 66
DB 10 KAGDLVFAKMGKYPHPWPARIDELPEGAVKPPANK-YPIFFGTHTETAFGLPKDLFPYKEY 68

QY 67 KEEMIKINKGRFOQAVDAVEEFLRRAGKDOTSSHNSSDDKRRNSSEERSRPN 121
DB 69 KDKFKSKNRKGFNEGLWEIENPGVKFTGYQAIQQSSSETEGGGNTADASS- 123

QY 122 SGDEKRLSLSEGKVKKNMGEGKRVSSGSRGSKSPKRAQEOSPRKRGPPPKDKOL 181
DB 124 -----EGDRVEEDGKGRKNEKAGSKRKSKSYTSKSKSKOSKSPG--DEDDKDC 171

QY 182 TTPES-STVKGMAG 195
DB 172 KEENKSSSEGGDAG 186

RESULT 12
Q8BPG7
ID Q8BPG7 PRELIMINARY; PRT; 237 AA.
AC Q8BPG7
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hepatoma-derived growth factor.
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Body;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,

RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.",
RL Nature 420:563-573(2002).
DR EMBL; AK076021; BAC36126.1; -;
SQ SEQUENCE 237 AA; 26270 MW; BBA0CF574DA4733F CRC64;

Query Match 14.1%; Score 186.5; DB 11; Length 237;
Best Local Similarity 29.8%; Pred. No. 1.5e-07;
Matches 65; Conservative 32; Mismatches 94; Indels 27; Gaps 9;

QY 7 RLGLDVLWGKLGKRYPPWPKGVNPPKDLKKPRGKCFVKKFFGTEDHAWIKVEQLKPYHAH 66
DB 11 KCGDLVFAKMGKYPHPWPARIDEMPEAAVKSTANK-YQVFFGTHTETAFGLPKDLFPYKEY 69

QY 67 KEEMIKINKGRFOQAVDAVEEFLRRAGKDOTSSHNSSDDKRRNSSEERSRPN 126
DB 70 KEKFGKPNKRGFSEGLWEIEN-----NPTVKASGYSSQKSCAAPEVEPEAHEGDDG 124

QY 127 RKLSEGLKVKKNMGEGKRVSSGSRGSKSPKLR-----AQEOSPRKRGPPK- 177
DB 125 KKGK-ABGSSDE---EGKLVIDEPAKEKNEKGTLLKRRAGDVLDSPKR-----PKESGDHE 176

QY 178 EKDTIPESSTVKGMAGPMAAFKQPTASEPVKDADP 215
DB 177 BED---KEIAALEGERPLPVEVEK-NSTPSEPDGSGGP 210

RESULT 13
Q8T079
ID Q8T079 PRELIMINARY; PRT; 602 AA.
AC Q8T079;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE LD22344p.
GN BEST:LD22483 OR BEST:LD29743 OR CG4747.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY069497; AAL39642.1; -;
DR FlyBase; FBGN0043456; BEST:LD22483.
DR InterPro; IPR006115; 6PGD_NAD.
DR InterPro; IPR000313; PWWP_domain.
DR Pfam; PF03446; NAD_binding_2; 1.
DR Pfam; PF00855; PWWP; 1.
DR SMART; SM00291; PWWP; 1.
DR PROSITE; PS0812; PWWP; 1.
SQ SEQUENCE 602 AA; 65252 MW; E004EEC610C23625 CRC64;

Query Match 14.1%; Score 186.5; DB 5; Length 602;
Best Local Similarity 25.5%; Pred. No. 4.3e-07;
Matches 61; Conservative 44; Mismatches 89; Indels 45; Gaps 10;

QY 10 DLVWGKLGKRYPPWPKGVNPPKDL---KKPRGKCFVKKFFGTEDHAWIKVEQLKPYHA- 65
DB 24 DLIAWAKGFTWPGMIVDPDLDSQQRRANTKCVF--PFGSRNFAMIEENIKPFEPG 81

QY 66 HKEEMIKINKGRFOQAVDAVEEFLRRAGKDOTSSHNSSDDKRRNSSEERSRPN 125
DB 82 WKEELAKVSKPAAPRAFRHANTIEKIYDDPAEVDQVNSCGAPNHTAEADFKIR--DGLD 139

```

Qy 126 KRKLSLSEGVKKVMGEGKRVVSGSSEB-----SKSPL-----KRAEQS 167
Db 140 SEEVGEEATADGNGVVAHV--GSPDEGLDVEINADSSASPVTSPAVTTKAGKRT 197
Qy 168 PR-----KRGPPKDEKDLTTPESSTVKGMMAGPMAAPKPOPTAS-EPVKDAD 214
Db 198 PKAKSVAATSVKSTGSAKQRTSAQOSP-----SGPSNAKRGKRDVSGEALQDAD 251

RESULT 14
Q9VL51 PRELIMINARY; PRT; 603 AA.
AC Q9VL51;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE CG4747 protein.
GN BEST:LD22483 OR CG4747.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA De Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foele C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Hernandez J.R., Houck J.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
[2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,

```

```

RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AE003627; AAF52846.2; -
DR FlyBase; FBgn0043456; BEST:LD22483.
DR InterPro; IPR000313; PWWP domain.
DR Pfam; PF00855; PWWP; 1.
DR SMART; SM00293; PWWP; 1.
DR PROSITE; PS00812; PWWP; 1.
SQ SEQUENCE 603 AA; 65354 MW; 98796823261F3D48 CRC64;

Query Match 14.1%; Score 186.5; DB 5; Length 603;
Best Local Similarity 25.5%; Pred. No. 4.3e-07;
Matches 61; Conservative 44; Mismatches 89; Indels 45; Gaps 10;

Qy 10 DLVWKLGRYPVPPGKIVNPDKL---KKPRGKKCFVKEFGTEHAWIKVQKPYHA- 65
Db 24 DLIAWAKMGKTPWPGMIVDPDLDDLSQRRANTKCVF--FFGRNFAWIBENNIKPEGP 81
Qy 66 HKEEMIKNGKRFQQAADVAVEFLRRAGKQDOTSSHNSDDDKNRNRSSBERRSPNCSDE 125
Db 82 WKELAKVSPAPAFRHWNTDIEKYDDPAEVDQVNVKSCGAPNHATEADFKIR--DGLD 139
Qy 126 KRKLSLSEGVKKVMGEGKRVVSGSSEB-----SKSPL-----KRAEQS 167
Db 140 SEEVGEEATADGNGVVAHV--GSPDEGLDVEINADSSASPVTSPAVTTKAGKRT 197
Qy 168 PR-----KRGPPKDEKDLTTPESSTVKGMMAGPMAAPKPOPTAS-EPVKDAD 214
Db 198 PKAKSVAATSVKSTGSAKQRTSAQOSP-----SGPSNAKRGKRDVSGEALQDAD 251

RESULT 15
Q8VHK7 PRELIMINARY; PRT; 237 AA.
AC Q8VHK7;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hepatoma-derived growth factor (fragment).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar;
RA Matsui H., Kawada N., Yokoya F., Takahara Y.;

```

0
9
9
2
. .
.
0
3
2
)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 8, 2003, 10:34:46 ; Search time 3727 Seconds
(without alignments)
2733.163 Million cell updates/sec

Title: US-09-987-755-2

Perfect score: 1319

Sequence: 1 MAASLRLGLDVGKGLRYP.....VCYQAITKKIKICEDLLPLR 249

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 577422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DBV=xlh

-MODE=frame+ p2n.model -DBV=xlh
-Q=/cgn2_1/USPTO.spool/US09987755/runat_081222003_103435_18308/app_query.fasta_1.391
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09987755_@CGN_1_1_3508_@runat_081222003_103435_18308 -NCFU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*

- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rod.*
- 36: em.htg.mam.*
- 37: em.htg.vrt.*
- 38: em.ev.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1291	97.9	1831	9	AF326966 Homo sapi
2	1291	97.9	3316	6	AX099496 Sequence
3	1291	97.9	3316	6	BD063979 Secreted
4	1247.5	94.6	3187	10	BC006893 Mus muscu
5	1138	86.3	2047	9	BC003693
6	915	69.4	140092	9	AL162815 Human DNA
7	915	69.4	162847	2	AL162815 Homo sapi
8	915	69.4	231260	2	AL160172 Homo sapi
9	595.5	45.1	658	6	BD059639 Secreted
10	551	41.8	154218	9	AC020663 Homo sapi
11	551	41.8	169166	9	AC027687 Homo sapi
12	514	39.0	233786	2	AC113755 Rattus no
13	514	39.0	237081	2	AC108242 Rattus no
14	514	39.0	243770	2	AC123492 Rattus no
15	505.5	38.3	181475	10	AC127246 Mus muscu
16	444	33.7	307	6	BD058512 Secreted
17	341	25.9	445	6	BD045202 Sequence
18	255.5	19.4	1893	3	AK114439 Ciona int
19	204.5	15.5	4487	6	AR220855 Sequence
20	201.5	15.3	1932	4	BT237996 Bos tauru
21	198.5	15.0	69061	2	AC012986 Drosophil
22	194.5	14.7	168469	3	AC007886 Drosophil
23	194.5	14.7	181056	3	AC009733 Drosophil
24	194.5	14.7	228443	3	AE003772 Drosophil
25	193.5	14.7	1423	9	AF110642 Homo sapi
26	193.5	14.7	1427	9	AB029156 Homo sapi
27	193.5	14.7	1817	9	BC015483 Homo sapi
28	193.5	14.7	1830	9	AF151900 Homo sapi
29	193.5	14.7	1973	6	BD156109 Primer fo
30	193.5	14.7	1973	9	AK001280 Homo sapi
31	193.5	14.7	3805	6	BD131023 Human gro
32	193	14.6	1859	3	AY051690 Drosophil
33	191.5	14.5	612	10	AB029493 Mus muscu
34	190.5	14.4	2750	10	AF389347 Rattus no
35	190	14.4	2123	3	AY069497 Drosophil
36	186.5	14.1	1744	10	BC005713 Mus muscu
37	186.5	14.1	1752	10	AF251787 Mus muscu
38	186.5	14.1	1763	10	BC021654 Mus muscu
39	185.5	14.1	714	6	E13960 Mouse mRNA
40	185.5	14.1	723	6	E08545 DNA encodin
41	185.5	14.1	1563	6	E13961 Mouse mRNA
42	185.5	14.1	1563	10	M0SHDGF
43	185.5	14.1	2096	9	BC018991 Homo sapi
44	185.5	14.1	2376	6	AR081931 Sequence
45	185.5	14.1	2376	6	AX302495 Sequence

ALIGNMENTS

RESULT 1

AF326966
LOCUS AF326966 1831 bp mRNA linear PRI 04-MAR-2001
DEFINITION Homo sapiens cytokine-like nuclear factor n-pac mRNA, complete cds.
ACCESSION AF326966
VERSION AF326966.1 GI:13194723
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1831)
New, L. and Han, J.
A novel cytokine-like nuclear factor, n-pac
Unpublished
REFERENCE 2 (bases 1 to 1831)
New, L. and Han, J.
Direct Submission
JOURNAL Submitted (07-DEC-2000) Immunology, The Scripps Research Institute,
10550 North Torrey Pines Rd, La Jolla, CA 92037, USA
FEATURES
source
1..1831
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1..1644
/codon_start=1
/product="cytokine-like nuclear factor n-pac"
/protein_id="AAK15524.1"
/db_xref="GI:13194724"
/translation="MAAVSLRLGLVWGKLRYPWPVKIYVNPVKDLKKRKKCFVF
KFFGTEHWIKVQLKPYHAHKEEMI KINKRFPQAVDAVEFLRAKGGKDTSSH
NSDDNRNRNSSEERSRNSGDEKRKLSLSEKVKNNKGEKRVSSGSSRRGSKSL
KRAOQSPKRPRPPKDEKDLTI PESSVTVMGMAGPMAAFKWPASEPVKADPHFH
HFLSOREKPAVCYQALTKLKI CEETSTGSIQADSTAVNGSITPTDKKIGFLGL
LMGSGVSNLNMKHTVTWRNRAKEGARLGRTPAEVWSTCDITFPACVDPKRAKOL
VLGSGVLQGIIRPKCYVDMSVTADVTTELAVQIVSRGGRFLPAVSGNQSLSDGM
LVILAAGRGYIEPSCSFQMGKTSFPLFVGNAAKMLIVNNVQGSFMTATIAEGLT
LAQVTSQSLQDLILNQGOLASIFLDQKCNILQGNPKDPFYLKIQKOLRLAIALG
DAVNHTPMAAAAEVYKRAKALDQSDNDMSAVRYVYIH"
BASE COUNT 487 a 472 c 510 g 362 t
ORIGIN
Alignment Scores:
Pred. No.: 7,26e-82 Length: 1831
Score: 1291.00 Matches: 243
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.59% Mismatches: 0
Query Match: 97.88% Indels: 0
DB: 9 Gaps: 0
US-09-987-755-2 (1-249) x AF326966 (1-1831)
QY 1 MetAlaAlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrPro 20
DB 1 ATGGCGGCTGTGAGTCTCGGCTCGGCGACTTGTGTGGGGGAAACTCGGCGGATATCCT 60
QY 21 ProTrpProGlyLysLeuValAsnProProLysAspLeuLysProArgGlyLysLys 40
DB 61 CTTTGGCCAGGAAAGATTGTTAATCCCAAGGACTTGAAGAAACCTCGCGGAAAGAAA 120
QY 41 CysPhePheValLysPheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeu 60
DB 121 TGCTTCTTTGTGAATTTTGGACACAGACATCATGCTGGATCAAAAGTGGACAGCTG 180
QY 61 LysProTyrHisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGln 80
DB 181 AAGCATATCATGCTCATAAAGAGAAATGATAAAATTAACAGGGTAAACGATTCCAG 240
QY 81 GlnAlaValAspAlaValGluGluPheLeuArgArgAlaLysGlyLysAspGlnThrSer 100
DB 241 CAAGCGGTAGATGCTGTGCAAGATTCTCTCAGGAGAGCCAAAGGAAAGACACACGTC 300
QY 101 SerHisAsnSerSerAspAspLysAsnArgArgAsnSerSerGluGluArgSerArgPro 120

Db 301 TCCACAAATCTTCTGATGACAGAAATCGAGTAATTCAGTGCAGGAGAGAAGTAGGCCA 360
QY 121 AsnSerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValLysValLysAsnMet 140
DB 361 AACTCAGGTGATGAGAAGCGCAAACTTAGCTGTCTGAAGGGAAGGTGAAGAAACATG 420
QY 141 GlyGluGlyLysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeu 160
DB 421 GGAGAAGGAAGAGAGAGGGTGTCTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTG 480
QY 161 LysArgAlaGlnGlnSerProArgLysArgGlyArgProProLysAspGlyLysAsp 180
DB 481 AAAAGAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTCCGCCCAAGAGATGAGAAGAT 540
QY 181 LeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPhe 200
DB 541 CTCACCATCCGGAGTCTAGTACCGTGAAGGGGATGATGCCGACCGATGCCCGCTTT 600
QY 201 LysTrpGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisPhe 220
DB 601 AAATGGCAGCCACCGCAAGCGAGCTGTTAAAGATGCAGATCCTCATTTCCATCATTC 660
QY 221 LeuLeuSerGlnThrGluLysProAlaValCysTyrGlnAlaIleThrLysLysLeuLys 240
DB 661 CTGCTAAGCCAAACAGAGAAGCCAGCTGTCTGTTCAGGCAATCAGCAAGAGTTCGAAA 720
QY 241 IleCysGluAsp 244
DB 721 ATATGTGAAGAG 732
RESULT 2
LOCUS AX099496 3316 bp DNA linear PAT 02-APR-2001
DEFINITION Sequence 136 from Patent WO0119988.
ACCESSION AX099496
VERSION AX099496.1 GI:13538581
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Jacobs, K., McCoy, J. M., Lavallie, E. R., Collins-Racie, L. A., Evans, C.,
Merberg, D., Treacy, M., Bowman, M. R., Spaulding, V. and Agostino, M. J.
TITLE Secreted proteins and polynucleotides encoding them
JOURNAL Patent: WO 0119988-A 136 22-MAR-2001;
Genetics Institute, Inc. (US)
FEATURES
source
1..3316
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 875 a 813 c 887 g 741 t
ORIGIN
Alignment Scores:
Pred. No.: 1.39e-81 Length: 3316
Score: 1291.00 Matches: 243
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.59% Mismatches: 0
Query Match: 97.88% Indels: 0
DB: 6 Gaps: 0
US-09-987-755-2 (1-249) x AX099496 (1-3316)
QY 1 MetAlaAlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrPro 20
DB 7 ATGGCGGCTGTGAGTCTCGGCTCGGCGACTTGTGTGGGGGAAACTCGGCGGATATCCT 66
QY 21 ProTrpProGlyLysLeuValAsnProProLysAspLeuLysProArgGlyLysLys 40
DB 67 CTTTGGCCAGGAAAGATTGTTAATCCCAAGGACTTGAAGAAACCTCGCGGAAAGAAA 126

Qy 41 CysPhePheValLysPhePheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeu 60
 Db 127 TGCTTCTTTGGAATTTTGGAAACAGAGATCATGCTGGATCAAAAGTGGACACCTG 186
 Qy 61 LysProThrHisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGln 80
 Db 187 AAGCCATATCATGCTCATAAAGAGAAATGATAAAATTAACAAAGGGTAAACGATTCAG 246
 Qy 81 GlnAlaValAspAlaValGluGluPheLeuArgAlaLysGlyLysAspGlnThrSer 100
 Db 247 CAACGGTAGATGCTGCGAAGATTCTCAGGAGAGCCAAAGGAAAGACACGACGTCA 306
 Qy 101 SerHisAsnSerSerAspLysAsnArgAsnSerSerGluGluArgSerArgPro 120
 Db 307 TCCACCAATCTCTGATGACAAGATCGAGTAAATCCAGTGAGGAGAGAGTAGGCCA 366
 Qy 121 AsnSerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMet 140
 Db 367 AACTCAGGTGATGAGAGCGCAAACTTAGCCGTGTCTGAAGGGAAAGGTGAAGAACATG 426
 Qy 141 GlyGluGlyLysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeu 160
 Db 427 GGAAGAGAAAGAGAGAGGGTGTCTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTG 486
 Qy 161 LysArgAlaGlnGluGlnSerProArgLysArgGlyArgProProLysAspGluLysAsp 180
 Db 487 AAAGAGCCCAAGAGCAAGTCCCGGAGCGGGTGGCCCGCCCAAGAGGATGAGAGAT 546
 Qy 181 LeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPhe 200
 Db 547 CTCACCATCCCGAGTCTAGTACCGTGAAGGGATGATGGCGGACCGATCGCCGCGCTT 606
 Qy 201 LysTrpGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisPhe 220
 Db 607 AAATGGCAGCCCAACCGCAAGCGCGCTGTAAAGATGAGATCTCATTTCCATCATTTTC 666
 Qy 221 LeuLeuSerGlnThrGluLysProAlaValCysTrpGlnAlaIleThrLysLysLeuLys 240
 Db 667 CTGCTAAGCCAAACAGAGAACCCAGCTGTCTGTACCAAGGCAATCAGAGAGATTGAAA 726
 Qy 241 IleCysGluAsp 244
 Db 727 ATATGTGAAGAG 738

RESULT 3
 BD063979
 LOCUS
 DEFINITION Secreted proteins and polynucleotides encoding them. PAT 27-AUG-2002
 ACCESSION BD063979
 VERSION BD063979.1 GI:22609582
 KEYWORDS JP 2001506848-A/3.
 SOURCE unidentified
 ORGANISM unclassified.
 REFERENCE 1 (bases 1 to 3316)
 AUTHORS Jacobs, K., McCoy, J.M., Lavallie, E.R., Racie, L.A., Merberg, D., Treacy, M., Spaulding, V. and Agostino, M.J.
 TITLE Secreted proteins and polynucleotides encoding them
 JOURNAL Patent: JP 2001506848-A 3 29-MAY-2001;
 GENETICS INSTITUTE INC
 COMMENT PN JP 2001506848-A/3
 PD 29-MAY-2001
 PF 12-DEC-1997 JP 1998525996
 PR 13-DEC-1996 US 08/766263
 PI KENNETH JACOBS, JOHN M MCCOY, EDWARD R LAVALLIE, LISA A RACIE, PI DAVID MERBERG.
 PI MAURICE TREACY, VIKKI SPAULDING, MICHAEL J AGOSTINO PC
 C12N15/12, C12N5/10, C07K14/47, A61K38/17
 CC Strandedness: Double;
 CC Topology: Linear;
 FH Key Location/Qualifiers.

FEATURES

source 1..3316
 /organism="unidentified"
 /mol_type="genomic DNA"
 /db_xref="taxon:32644"
 BASE COUNT 875 a 813 c 887 g 741 t
 ORIGIN

Alignment Scores:

Pred. No.: 1..39e-81 Length: 3316
 Score: 1291.00 Matches: 243
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 99.59% Mismatches: 0
 Query Match: 97.88% Indels: 0
 DB: 6 Gaps: 0

US-09-987-755-2 (1-249) x BD063979 (1-3316)

Qy 1 MetAlaAlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTrpPro 20
 Db 7 ATGGCGCTGTGAGTCTGCGGCTCGCGACTTGGTGTGGGGGAAACTCGGCCGATATCTCT 66
 Qy 21 ProTrpProGlyLysIleValAsnProProLysAspLeuLysLysProArgGlyLysLys 40
 Db 67 CCTTGGCCAGGAAGATTGTTATCCCAAGACTTGAAGAACTTCGCGGAAGAAA 126
 Qy 41 CysPhePheValLysPhePheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeu 60
 Db 127 TGCTTCTTTGGAATTTTGGAAACAGAGATCATGCTGGATCAAAAGTGGACACCTG 186
 Qy 61 LysProThrHisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGln 80
 Db 187 AAGCCATATCATGCTCATAAAGAGAGAAATGATAAAATTAACAAAGGGTAAACGATTCAG 246
 Qy 81 GlnAlaValAspAlaValGluGluPheLeuArgAlaLysGlyLysAspGlnThrSer 100
 Db 247 CAACGGTAGATGCTGCGAAGATTCTCAGGAGAGCCAAAGGAAAGACACGACGTCA 306
 Qy 101 SerHisAsnSerSerAspLysAsnArgAsnSerSerGluGluArgSerArgPro 120
 Db 307 TCCACCAATCTCTGATGACAAGATCGAGTAAATCCAGTGAGGAGAGAGTAGGCCA 366
 Qy 121 AsnSerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMet 140
 Db 367 AACTCAGGTGATGAGAGCGCAAACTTAGCCGTGTCTGAAGGGAGGTGAAGAACATG 426
 Qy 141 GlyGluGlyLysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeu 160
 Db 427 GGAAGAGAAAGAGAGGGTGTCTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTG 486
 Qy 161 LysArgAlaGlnGluGlnSerProArgLysArgGlyArgProProLysAspGluLysAsp 180
 Db 487 AAAGAGCCCAAGAGCAAGTCCCGGAGCGGGTGGCCCGCCCAAGAGGATGAGAGAT 546
 Qy 181 LeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPhe 200
 Db 547 CTCACCATCCCGAGTCTAGTACCGTGAAGGGATGATGGCGGACCGATCGCCGCGCTT 606
 Qy 201 LysTrpGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisPhe 220
 Db 607 AAATGGCAGCCCAACCGCAAGCGCGCTGTAAAGATGAGATCTCATTTCCATCATTTTC 666
 Qy 221 LeuLeuSerGlnThrGluLysProAlaValCysTrpGlnAlaIleThrLysLysLeuLys 240
 Db 667 CTGCTAAGCCAAACAGAGAACCCAGCTGTCTGTACCAAGGCAATCAGAGAGATTGAAA 726
 Qy 241 IleCysGluAsp 244
 Db 727 ATATGTGAAGAG 738

RESULT 4
 BC006893
 LOCUS
 DEFINITION Mus musculus RIKEN cdna 3930401K13 gene, mRNA (cdna clone MGC:11922

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 140092)
Heath, P.
Direct Submission
Submitted (08-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jul 28, 2000 this sequence version replaced gi:9408285.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission only a small overlap as described above.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, ENBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/projects/c_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr20>
IMPORTANT: This sequence is not the entire insert of clone RPS-1041C10 it may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.
The true left end of clone RPS-963K23 is at 139993 in this sequence. The true right end of clone RPS-1063B2 is at 100 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one ML3 subclone; and the assembly was confirmed by restriction digest. RPS-1041C10 is from the library RPS-5 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>
VECTOR: pCVPAC2.

FEATURES
source
1..140092
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="20"
/clone="RPS-1041C10"
/clone_lib="RPS-5"
1229..11513
/note="AluJb repeat: matches 1..286 of consensus"
repeat_region
1548..1870
/note="AluX repeat: matches 1..306 of consensus"
repeat_region
2266..2548
/note="AluSg repeat: matches 1..297 of consensus"
repeat_region
3090..3384
/note="AluSg repeat: matches 1..299 of consensus"
repeat_region
3995..4317
/note="AluSg repeat: matches 1..310 of consensus"
repeat_region
4483..4694
/note="AluJb repeat: matches 87..300 of consensus"
repeat_region
5798..5993
/note="MIR repeat: matches 6..229 of consensus"
repeat_region
6670..6981
/note="156 copies 2 mer gg 54% conserved"
misc_feature
6709..7745
/note="CpG island"
/evidence=not_experimental
complement(71176..7401)
/gene="B4GALT5"
/evidence=not_experimental
complement(<71176..7401)
/gene="B4GALT5"
/product="dJ1041C10.1 (beta-1,4-galactosyltransferase, polypeptide 5)"

CDS

/note="match: cDNAs: Em:AF142673 Em:AF038663 Em:AF055478 Em:Z71594 Em:X78947 Em:AB004550
match: ESTs: Em:AI017532 Em:AI059051 Em:AI095142 Em:AI425818 Em:AI029693 Em:AI151021 Em:AW122637 Em:AI618828 Em:AI088794 Em:AA956979"
/evidence=not_experimental
complement(<71176..7290)
/gene="B4GALT5"
/note="continues in dJ1063B2 (AL035683)
match: proteins: Sw:P17135 Tr:O60514 Tr:Q9YAN1 Tr:Q9QY11 Tr:Q9TU66 Sw:P17136 Tr:Q9WVK5 Sw:P14648 Tr:O88419 Sw:P14678"
/codon_start=1
/evidence=not_experimental
/product="dJ1041C10.1 (beta-1,4-galactosyltransferase, polypeptide 5)"
/protein_id="CAC11145.1"
/db_xref="GI:10638624"
/translation="MRARRGLRLPRSLAALFFSLSSLLYFVYVAPGI"
7256..7420
/note="55 copies 3 mer cgc 60% conserved"
7860..8312
/note="L2 repeat: matches 2252..2750 of consensus"
8399..8493
/note="MIR repeat: matches 90..185 of consensus"
8931..9113
/note="MIR repeat: matches 28..227 of consensus"
9182..9376
/note="MIR repeat: matches 21..219 of consensus"
9460..9616
/note="L1MC4 repeat: matches 7801..7968 of consensus"
9617..9913
/note="AluJb repeat: matches 1..301 of consensus"
9914..10041
/note="L1MC4 repeat: matches 7677..7801 of consensus"
10170..10483
/note="AluSg repeat: matches 1..304 of consensus"
10484..10650
/note="AluSg/x repeat: matches 130..297 of consensus"
AluSg/x repeat: matches 130..297 of consensus"
10661..10961
/note="AluSg repeat: matches 2..304 of consensus"
11129..11429
/note="L2 repeat: matches 2436..2744 of consensus"
11743..11950
/note="MIR repeat: matches 23..245 of consensus"
12144..12452
/note="AluSg repeat: matches 1..308 of consensus"
12476..12622
/note="MLT11 repeat: matches 201..368 of consensus"
12623..12914
/note="AluSx repeat: matches 1..293 of consensus"
12915..12958
/note="MLT11 repeat: matches 368..407 of consensus"
12984..13142
/note="L1ME3 repeat: matches 5725..5882 of consensus"
13153..13356
/note="MER46C repeat: matches 3..193 of consensus"
MER46C repeat: matches 3..193 of consensus"
13371..13554
/note="AluJb repeat: matches 128..309 of consensus"
13637..13940
/note="AluSg repeat: matches 1..305 of consensus"
13943..14011
/note="MER46C repeat: matches 268..335 of consensus"
MER46C repeat: matches 268..335 of consensus"
14116..14177
/note="MLT1G repeat: matches 67..127 of consensus"
MLT1G repeat: matches 67..127 of consensus"
14178..14502
/note="AluSx repeat: matches 1..312 of consensus"
14503..14880
/note="MLT1G repeat: matches 127..512 of consensus"

```

MLRIG repeat: matches 127. .512 of consensus"
14887. .15197
/note="Alusx repeat: matches 1. .302 of consensus"
15314. .15370
/note="L1ME3 repeat: matches 5527. .5583 of consensus"
L1MC/D repeat: matches 5527. .5583 of consensus"
15349. .15874
/note="7SLRNA repeat: matches 1. .319 of consensus"
7SLRNA repeat: matches 1. .319 of consensus"
15884. .16168
/note="Alusq repeat: matches 3. .313 of consensus"
16317. .16632
/note="Alusx repeat: matches 1. .297 of consensus"
complement(16709. .17162)
/note="match: GSS: Em:AQ628320"
17164. .17592
/note="match: GSS: Em:AQ469272"
17231. .17359
/note="L1ME3 repeat: matches 6042. .6164 of consensus"
L1ME3 repeat: matches 6042. .6164 of consensus"
17435. .17528
/note="MER47A repeat: matches 4. .99 of consensus"
MER47A repeat: matches 4. .99 of consensus"
17529. .17835
/note="Alusx repeat: matches 1. .307 of consensus"
17836. .18113
/note="MER47A repeat: matches 99. .366 of consensus"
MER47A repeat: matches 99. .366 of consensus"
18116. .18317
/note="L1ME3 repeat: matches 5697. .5898 of consensus"
L1ME3 repeat: matches 5697. .5898 of consensus"
18981. .19288
/note="Alusx repeat: matches 1. .305 of consensus"
19723. .19853
/note="FLAMC repeat: matches 19. .121 of consensus"
19876. .20178
/note="Alusx repeat: matches 1. .304 of consensus"
20203. .20241
/note="MIR repeat: matches 203. .241 of consensus"
20255. .20553
/note="Alusq repeat: matches 1. .299 of consensus"

Alignment Scores:
Pred. No.: 2.15e-53 Length: 140092
Score: 915.00 Matches: 204
Percent Similarity: 53.04% Conservative: 16
Best Local Similarity: 58.45% Mismatches: 24
Query Match: 69.37% Indels: 108
DB: 9 Gaps: 1

US-09-987-755-2 (1-249) x AL162615 (1-140092)

Qy 1 MetAlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyPro 20
Db 122725 ATGGGGCTGTGAGTCTCGCTCAGCAACTGTGATGGGGAACTCGGCTGTATCCT 122784
Qy 21 ProTrpProGlyLysLeuValAsnProProLysAspLeuLysProArg----- 37
Db 122785 CCTTACCAGGAAGATCGTTAATCCACCGAAGACTTTGAAGAAACCTCA-TGGGGCTGG 122843
Qy 37 ----- 37
Db 122844 GCGGGTGGCTCAGCCCTGTGATCACAGCACTTTGGGAGGCCAGGTGGGGGATCACA 122903
Qy 37 ----- 37
Db 122904 GGTCAGGAGATTGAGACCATCTCGCTAACATGGTGAATCCCGTCTATATACTAAAAATAC 122963
Qy 37 ----- 37
Db 122964 AAAAAAATAAATTAGCAGCGGTGGTGGCGGCGCTTAGTCTCTAGCTCTCAGGAGGCT 123023
Qy 37 ----- 37

```

```

Db 123024 GAGGAGGAGAGATAGCATGAACCCGGGAGGTGGAGCTTGAGTGCAGTGCAGCCGAGATGCGACCA 123083
Qy 37 ----- 37
Db 123084 CTGCACCTCAACCTGGGCAACAGTGCAGACTCCATCTCAAAAAAAGAAACCTCAT 123143
Qy 38 GlyLysCysPhePheValLys-PhePheGlyThrGluAspHisAlaTrpLysVal 57
Db 123144 GGAAGAAATGCTCTCTTTGTGAAATTTTTTTTGGAAACATAAGATCATCGCTGGATTAAGT 123203
Qy 57 lGluGlnLeuLysProTyRHisAlaHisLysGluGluMetLleLysLleAsnLysGlyLys 77
Db 123204 GAAACAGCTGAAGCCATATCATCTCATAAAGAGGAATGATAAAGATTCAAGAGGTAA 123263
Qy 77 sArgPheGlnGlnAlaValAspAlaValGluGluPheLeuArgArgAla-LysGlyLysA 97
Db 123264 ACTATTCCAGCAAGCGGTGATGCCCAAGAGATTCTTCAGAGAGCCCAAGAGGAAG 123323
Qy 97 spGlnThrSerSerHisAsnSerSerAspLysAsnArgArgAsnSerSerGluGluA 117
Db 123324 ATCAGAGCTCATCCCAATTCCTGTGATGACAAGAAATCGCTTAATTCCAGTGAGAGA 123383
Qy 117 rSerArgProAsnSerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValL 137
Db 123384 GAAAGCAGGCCAAACTCAGGTGATGAGAAGTGCAAC--AGCCTGTCTTGAAGGAAAGTGA 123441
Qy 137 yLysAsnMetGlyGluGlyLysValArgValSerSerGlySerSerGluArgGlySerL 157
Db 123442 AGAAGATCATGGGAGATGGAAGAGGGGTGTCTTCGGGCTCTTCAGAGACAGGCTCCA 123501
Qy 157 ySerProLeuLysArgAlaGlnGluGlnSerProArgLysArgGlyArgProLysA 177
Db 123502 AATCCCTCTGAAAGAAATCCAGAGCAAGATGCCAGAGCGGGTGCAGCCCCCAAGA 123561
Qy 177 spGluLysAspLeuThrLleProGluSerSerThrValLysGlyMetMetAlaGlyProM 197
Db 123562 ATGAGAAGAAATGTCCCATCGTGAGTCCGGTACTTCTTGAAGTGGATGATGCCGAGCAG 123621
Qy 197 etAlaAlaPheLysTrpGlnProThrAlaSerGluProValLysAspAlaAspProHisP 217
Db 123622 TGGCTGATTTAAATGGCAGCCAGAGCGGCTGTTAAGATGCAATCTCAT 123681
Qy 217 heHisHisPheLeuSerGlnThrGluLysProAlaValCysTyrGlnAlaLleThrL 237
Db 123682 TCCATCATTTCTGCTAAGCCAAACAGCCATC--TCTGTCTGTACCAGGCAATCACA 123739
Qy 237 yLysLeuLysLleCysGluAsp 244
Db 123740 AGAAGTTGAAATATATGTGAAGAG 123762

RESULT 7
AC025980 162847 bp DNA linear HTG 07-JUN-2000
LOCUS Homo sapiens chromosome 20 clone RP11-444K9 map 20, WORKING DRAFT
DEFINITION SEQUENCE, 30 unordered pieces.
AC025980
AC025980.3 GI:8317124
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 162847)
BIRREN,B., LINTON,L., NUSBAUM,C., LANDER,E., ABRAHAM,H., ALLEN,N.,
ANDERSON,S., BALDWIN,J., BARNA,N., BASTIEN,V., BEDA,F.,
BOGUSLAVSKY,L., BOUKHALTER,B., BROWN,A., BURKETT,G.,
CAMPOLANO,A., CASTLE,A., CHOPEL,Y., COLANGELO,M., COLLINS,S.,
COLLYMORE,A., COOKE,P., DEARELLANO,K., DEWAR,K., DIAZ,J.S.,
BIRREN,B., LINTON,L., NUSBAUM,C., LANDER,E., ABRAHAM,H., ALLEN,N.,
ANDERSON,S., BALDWIN,J., BARNA,N., BASTIEN,V., BEDA,F.,
BOGUSLAVSKY,L., BOUKHALTER,B., BROWN,A., BURKETT,G.,
CAMPOLANO,A., CASTLE,A., CHOPEL,Y., COLANGELO,M., COLLINS,S.,
COLLYMORE,A., COOKE,P., DEARELLANO,K., DEWAR,K., DIAZ,J.S.,

```

Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczyk,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheters,R., Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivier,T.M., Oliver,J., Peterson,K., Pierre,N., Pisanic,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission

Submitted (18-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jun 7, 2000 this sequence version replaced gi:7652052.

All repeats were identified using RepeatMasker:

Smit, A.P.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE JOURNAL

COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L6971

Center clone name: 444_K9

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 150127 bases at least Q40

Consensus quality: 156059 bases at least Q30

Consensus quality: 158474 bases at least Q20

Insert size: 176000; agarose-fp

Insert size: 159947; sum-of-contigs

Quality coverage: 4.2 in Q20 bases; agarose-fp

Quality coverage: 4.6 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 30 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 1179: contig of 1179 bp in length

* 1180 1279: gap of 100 bp

* 1280 2395: contig of 1116 bp in length

* 2396 2495: gap of 100 bp

* 2496 3581: contig of 1086 bp in length

* 3582 3681: gap of 100 bp

* 3682 4980: contig of 1299 bp in length

* 4981 5080: gap of 100 bp

* 5081 6590: contig of 1510 bp in length

* 6591 6690: gap of 100 bp

* 6691 8194: contig of 1504 bp in length

* 8195 8294: gap of 100 bp

* 8295 9957: contig of 1653 bp in length

* 9958 10057: gap of 100 bp

* 10058 11940: contig of 1883 bp in length

* 11941 12040: gap of 100 bp

* 12041 13442: contig of 1402 bp in length

* 13443 13542: gap of 100 bp

* 13543 16309: contig of 2767 bp in length

* 16310 16409: gap of 100 bp

* 16410 18496: contig of 2087 bp in length

* 18497 18596: gap of 100 bp

* 18597 21162: contig of 2566 bp in length

* 21163 21262: gap of 100 bp

* 21263 24271: contig of 3009 bp in length

* 24272 24371: gap of 100 bp

* 24372 26834: contig of 2463 bp in length

* 26835 26934: gap of 100 bp

* 26935 29054: contig of 2120 bp in length

* 29055 29154: gap of 100 bp

* 29155 32187: contig of 3033 bp in length

* 32188 32287: gap of 100 bp

* 32288 35290: contig of 3003 bp in length

* 35291 35390: gap of 100 bp

* 35391 37991: contig of 2601 bp in length

* 37992 38091: gap of 100 bp

* 38092 41561: contig of 3470 bp in length

* 41562 41661: gap of 100 bp

* 41662 48058: contig of 6397 bp in length

* 48059 48158: gap of 100 bp

* 48159 52403: contig of 4245 bp in length

* 52404 52503: gap of 100 bp

* 52504 58539: contig of 6036 bp in length

* 58540 58639: gap of 100 bp

* 58640 64768: contig of 6129 bp in length

* 64769 64868: gap of 100 bp

* 64869 74871: contig of 10003 bp in length

* 74872 74971: gap of 100 bp

* 74972 85271: contig of 10300 bp in length

* 85272 85371: gap of 100 bp

* 85372 94554: contig of 9183 bp in length

* 94555 94654: gap of 100 bp

* 94655 102707: contig of 8053 bp in length

* 102708 102807: gap of 100 bp

* 102808 116262: contig of 13455 bp in length

* 116263 116362: gap of 100 bp

* 116363 134631: contig of 18269 bp in length

* 134632 134731: gap of 100 bp

* 134732 162847: contig of 28116 bp in length.

FEATURES

source

1. .162847
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="20"
/map="20"
/clone="RP11-444K9"
/CloneLib="RPC1-11 Human Male BAC"

misc_feature

1. .1179

/note="assembly_fragment"

1280. .2395

/note="assembly_fragment"

2496. .3581

/note="assembly_fragment"

3682. .4980

/note="assembly_fragment"

5081. .6590

/note="assembly_fragment"

6691. .8194

/note="assembly_fragment"

8295. .9957

/note="assembly_fragment"

10058. .11940

/note="assembly_fragment"

12041. .13442

/note="assembly_fragment"

13543. .16309

/note="assembly_fragment"

16410. .18496

/note="assembly_fragment"

18597. .21162

/note="assembly_fragment"

21263. .24271

/note="assembly_fragment"

24372. .26834

/note="assembly_fragment"

24372. .26834

/note="assembly_fragment"

24372. .26834

/note="assembly_fragment"

24372. .26834

/note="assembly_fragment"

24372. .26834

/note="assembly_fragment"


```

misc_feature /note="assembly_fragment"
26935..29054
/note="assembly_fragment"
clone_end:T7
vector_side:right"
29155..32187
/note="assembly_fragment"
32288..35290
/note="assembly_fragment"
35391..37991
/note="assembly_fragment"
38092..41561
/note="assembly_fragment"
41662..48058
/note="assembly_fragment"
48159..52403
/note="assembly_fragment"
52504..58539
/note="assembly_fragment"
58640..64768
/note="assembly_fragment"
64869..74871
/note="assembly_fragment"
74972..85271
/note="assembly_fragment"
85372..94554
/note="assembly_fragment"
94655..102707
/note="assembly_fragment"
clone_end:SP6
vector_side:right"
102808..116262
/note="assembly_fragment"
116363..134631

```

Alignment Scores:

```

Pred. No.: 2,53e-53 Length: 162847
Score: 915.00 Matches: 204
Percent Similarity: 63.04% Conservative: 16
Best Local Similarity: 58.45% Mismatches: 24
Query Match: 69.37% Indels: 108
DB: 2 Gaps: 1

```

US-09-987-755-2 (1-249) x AC025980 (1-162847)

```

Qy 1 MetAlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrPro 20
Db 24776 ATGGGGCTGTGAGTTCGCTGCTCAGCAACTTGATGGGGGAAACTCGGCTGTATCCT 24835

Qy 21 ProTribProGlyLysLeuValAsnProProLysAspLeuLysLysProArg----- 37
Db 24836 CCTTACCAGGAAGATCGTTATCCACCGAGGACTTGAAGAAACCTCA-TGGGGCTGG 24894

Qy 37 ----- 37
Db 24895 GCGGGTGGCTCAGCGCTGTGATCACAGCACTTTGGGAGGCGCAAGGTGGGGGATCACGA 24954

Qy 37 ----- 37
Db 24955 GGTACGAGATGAGACCATCTGGCTTAACATGGTGAATCCCGTCTATATAAATAC 25014

Qy 37 ----- 37
Db 25015 AAAAAAAAAATTAGCCAGCGGTGGTGGGGGGCTAGTCTAGCTCTCAGGAGGCT 25074

Qy 37 ----- 37
Db 25075 GAGGAGGAGATAGCATGAACCGGAGGTGGAGCTTGAGTGGCGGAGATCGCACCA 25134

Qy 37 ----- 37
Db 25135 CTGCACTCCAACCTGGGCAACAGAGTGAGACTCCATCTCAAAAAAAAAAAGAACCTCAT 25194

```

```

Qy 38 GlyLysLysCysPhePheValLys-PhePheGlyThrGluAspHisAlaTrpIleLysVa 57
Db 25195 GGAAGAAATGCTTCTTCTGTAATTTTTTTTGGACATAGATCATGCTCGGATTAAGT 25254

Qy 57 lGluGlnLeuLysProTyrHisAlaHisLysGluGluMetIleLysIleAsnLysGlyLy 77
Db 25255 GAAACAGCTGAAGCCATATCATGCTCATAAAGAGGAATGATAAAGATTTCACAGGTAA 25314

Qy 77 sArgPheGlnGlnAlaValAspAlaValGluGluPheLeuArgArgAla-LysGlyLysA 97
Db 25315 ACTATTCCAGCAAGCGGTGATGCCCAAGAGTTTCTCAGGAGAGCCCAAGGAAAG 25374

Qy 97 spGlnThrSerHisAenSerSerAspAspLysAsnArgArgAenSerSerGluGluA 117
Db 25375 ATCAGACGTCATCCCAATTTCTGTGATGACAAAGAAATCGGCTTAATTCAGTGAGGAGA 25434

Qy 117 xgSerArgProAenSerGlyAspGluLysArgLysLeuSerLeuGlyLysValL 137
Db 25435 GAAGCAGGCCAACTCAGGTGATGAGAAAGTGCAAAC--AGCCTGTCTGAAGGAAGATGA 25492

Qy 137 yLysAsnMetGlyGluGlyLysLysArgValSerSerGlySerSerGluArgGlySerL 157
Db 25493 AGAAGATCATGGAGATGGAAGAGGGGTGTCTTCGGGCTCTTCAGAGACAGGCTCCA 25552

Qy 157 ySerProLysArgAlaGlnGlnSerProArgLysArgGlyArgProLysA 177
Db 25553 AATCCCTCTGAAAAGAAATCAAGAGCAAAAGTCCCAAGAGGGGTGACGCCCAAGA 25612

Qy 177 spGluLysAspLeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProM 197
Db 25613 ATGAGAAGAAATGTCACCATCGTGAGTCCGCTGACCTTGAAGTGGATGGCGGAGCCAG 25672

Qy 197 etAlaAlaPheLysTrpGlnProThrAlaSerGluProValLysAspAlaAspProHisP 217
Db 25673 TGGCTCATTTAAATGCGAGCCGACCAAGAGCGGCTGTAAAGATGAGATCCTCAT 25732

Qy 217 heHisHisPheLeuSerGlnThrGluLysPProAlaValCysTyrGlnAlaIleThrL 237
Db 25733 TCCATCATTTCTGCTAAGCAAAACAAAGCATC--TCTGTCTGTACCAGGCAATCACAA 25790

Qy 237 yLysLeuLysLysCysGluAsp 244
Db 25791 AGAAGTTGAAAATATGTGAAGAG 25813

RESULT 8
AL160172 231260 bp DNA linear HTG 13-JUN-2001
LOCUS Homo sapiens chromosome 1 clone RP11-6302 map q25.2-31.1, ***
DEFINITION SEQUENCING IN PROGRESS ***, 9 unordered pieces.
ACCESSION AL160172 GI:14456182
VERSION HTG; HTGS_PHASE1; HTGS_CANCELLED.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Plumb, B.
TITLE Direct Submission
JOURNAL Submitted (13-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jun 14, 2001 this sequence version replaced gi:13373898.
COMMENT ----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA6302
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: M13; M7815; 1% of reads

```

Sequencing vector: plasmid; L08752; 98% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Consensus quality: 229290 bases at least Q40
 Consensus quality: 229747 bases at least Q30
 Consensus quality: 230088 bases at least Q20
 Inset size: 230460; sum-of-contigs
 Inset size: 235881; 6.2% error; agarose-fp
 Quality coverage: 8.80x in Q20 bases; sum-of-contigs Quality
 coverage: 9.32x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 9 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1 32893: contig of 32893 bp in length
* 32894 32993: gap of 100 bp
* 32994 48692: contig of 15699 bp in length
* 48693 48792: gap of 100 bp
* 48793 64328: contig of 15536 bp in length
* 64329 64428: gap of 100 bp
* 64429 106773: contig of 42345 bp in length
* 106774 106873: gap of 100 bp
* 106874 120306: contig of 13433 bp in length
* 120307 120406: gap of 100 bp
* 120407 167991: contig of 47585 bp in length
* 167992 168091: gap of 100 bp
* 168092 176543: contig of 8452 bp in length
* 176544 176643: gap of 100 bp
* 176644 208941: contig of 32198 bp in length
* 208942 208941: gap of 100 bp
* 208942 231260: contig of 22319 bp in length.

```

FEATURES

source

```

1..231260
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/map="q25.2-31.1"
/clone="RP11-6302"
/clone_lib="RPC1-11.1"

```

misc_feature

```

1..32893
/note="assembly_fragment:02273"

```

```

clone_end:SP6
vector_side:left

```

```

32994..48692
/note="assembly_fragment:04073"

```

```

fragment_chain:1

```

```

48793..64328
/note="assembly_fragment:03251"

```

```

fragment_chain:1

```

```

64429..106773
/note="assembly_fragment:02028"

```

```

fragment_chain:1

```

```

106874..120306
/note="assembly_fragment:01335"

```

```

fragment_chain:1

```

```

120407..167991
/note="assembly_fragment:01561"

```

```

fragment_chain:2

```

```

168092..176543
/note="assembly_fragment:04212"

```

```

fragment_chain:2

```

```

176644..208841
/note="assembly_fragment:01546.0"

```

```

208942..231260
/note="assembly_fragment:03725"

```

BASE COUNT 61279 a 52930 c 52964 g 63281 t 806 others
 ORIGIN

Alignment Scores:

```

Pred. No.: 3.72e-53 Length: 231260
Score: 915.00 Matches: 204
Percent Similarity: 63.04% Conservative: 16
Best Local Similarity: 58.45% Mismatches: 24
Query Match: 69.37% Indels: 108
DB: 2 Gaps: 1
US-09-987-755-2 (1-249) x AL160172 (1-231260)

Qy 1 MetlalaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyPro 20
Db 66485 ATGGGGCTGTGAGTTTCGCTGCTCAGCAACTGATGTGGGGAACTCGGCTGTATCCT 66544
Qy 21 ProTrpProGlyLysIleValAsnProProLysAspLeuLysLysProArg----- 37
Db 66545 CCTTCACAGGAAGATCGTTTATCCACCGNAGACTTGAAGAACTTCA-TGGGGGTGG 66603
Qy 37 ----- 37
Db 66604 GCGGGTGGCTCAGCGCTGTGTGATCACAGCACTTTGGGAGGCCAAGGTGGGGGATCACGA 66663
Qy 37 ----- 37
Db 66664 GGTCAAGAGATTGAGACCATTCTGGCTAACATGGTGAATCCCGTCTATATAAAATAC 66723
Qy 37 ----- 37
Db 66724 AAAAAAAAAATTAGCCAGCGGTGGTGGGGGGCTTAGTCTCTAGTCTCTCAGAGGCT 66783
Qy 37 ----- 37
Db 66784 GAGCAGGAGATAGCATGAACCGGGAGGTGGAGCTTGCAAGTGCAGCGAGATCGCACCA 66843
Qy 37 ----- 37
Db 66844 CTGCACCTCAACCTGGGCAACAGAGTGAGATCCATCTCAAAAAAAGAACCTCAT 66903
Qy 38 GlyLysLysCysPhePheValLys-PhePheGlyThrGluAspHisAlaTrpIleLysVa 57
Db 66904 GGAAGAAGATGCTTCTTTTGAAATTTTTTTGGAAACATAAGATCATGCTCGTGGATTAAGT 66963
Qy 57 lGluGlnLeuLysProTyrHisAlaHisLysGluGluMetIleLysIleAsnLysGlyLy 77
Db 66964 GAAACAGCTGAAGCCATATCATGCTCAATAAGAGAAATGATAAAGATTCAAGGGTAA 67023
Qy 77 sArgPheGlnGlnAlaValAspAlaValGluGluPheLeuArgArgAla-LysGlyLysA 97
Db 67024 ACTATTCAGCAAGCGGTGGATGCCCAAGAGATTTTCTCAGAGAGAGCCCAAGGGAAG 67083
Qy 97 spGlnThrSerSerHisAsnSerSerAspAspLysAsnArgArgAsnSerSerGluGluA 117
Db 67084 ATCAGAGTCATCCCAATTTCTGCTGATGACAGAAATCGGCTTAATTCAGTGAGGAGA 67143
Qy 117 rgSerArgProAsnSerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValL 137
Db 67144 GAACAGCGCCAAACTCAGGTGATGAGAAGTGCAAC--AGCCTGCTCGAAGAAAGTGA 67201
Qy 137 YsLysAsnMetGlyGluGlyLysLysArgValSerSerGlySerSerGluArgGlySerL 157
Db 67202 AGAAGATCATGGGAGTGGAAAGAGGGGGTGTCTTCGGGCTCTTCAGAGACAGGCTCCA 67261
Qy 157 YsSerProLeuLysArgAlaGlnGlnSerProArgLysArgGlyArgProProLysA 177
Db 67262 AATCCCTCTGAAAGAAATCCAGAGCAAGATGCCAGAGCGGGGTTCAGCCCCCAAGA 67321
Qy 177 spGluLysAspLeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProm 197
Db 67322 ATGAGAAGATGTCCACCATCGTGGAGTCCGCTACCTTGAAGTGGATGATGATGCCCGGAC 67381
Qy 197 etAlaAlaPheLysTrpGlnProThrAlaSerGluProValLysAspAlaAspProHisP 217
Db 67382 TGGCTGCAATTTAAATGGCAGCCGACCAAGAGCGGCGCTGTTAAAGATGCAGATCTCAT 67441

```

```

QY 217 hehisheLeuLeuSerGlnThrGluLysProAlaValCysTyrGlnAlaIleThrL 237
Db 67442 TCCATCATTTCTGCTAAGCCAAACAGCCATC-TCTGTCGTGTACCAGGCAATCACAA 67499
QY 237 yslsLeuLysIleCysGluasp 244
Db 67500 AGAAGTTGAAATATGTGAAGAG 67522

RESULT 9
BD059639/c
LOCUS BD059639 658 bp DNA linear PAT 27-AUG-2002
DEFINITION Secreted expressed sequence tags (ESTs).
ACCESSION BD059639
VERSION BD059639.1 GI:22605245
KEYWORDS JP 2001519666-A/1494.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 658)
AUTHORS Jacobs, K., McCoy, J.M., Lavallie, E.R., Racie, L.A., Merberg, D.,
Trecy, M., Spaulding, V. and Agostino, M.J.
TITLE Secreted expressed sequence tags (ESTs)
JOURNAL Patent: JP 2001519666-A 1494 23-OCT-2001;
GENETICS INSTITUTE INC
COMMENT PN JP 2001519666-A/1494
PD 23-OCT-2001
PF 10-APR-1998 JP 1998543068
PR 10-APR-1997 US 08/835913
PI KENNETH JACOBS, JOHN M MCCOY, EDWARD R LAVALLIE, LISA A RACIE, PI
DAVID MERBERG,
PI MAURICE TREACY, VIKKI SPAULDING, MICHAEL J AGOSTINO PC
C12N15/12, C12N5/10, C07K14/47, C12Q1/68, A61K38/17 CC Strandedness:
Double;
CC Topology: Linear;
FH Key Location/Qualifiers.
FEATURES
source
1..658
/mol_type="genomic DNA"
/db_xref="taxon:4577"
BASE COUNT 148 a 175 c 130 g 205 t
ORIGIN

Alignment Scores:
Pred. No.: 1.65e-33 Length: 658
Score: 595.50 Matches: 119
Percent Similarity: 97.54% Conservative: 0
Best Local Similarity: 97.54% Mismatches: 2
Query Match: 45.15% Indels: 1
DB: 6 Gaps: 1

US-09-987-755-2 (1-249) x BD059639 (1-658)

QY 91 ArgArgAlaLysGlyLysAspGlnThrSerSerHisAsnSerSerAspLysAsnArg 110
Db 375 AGGAGACCCAGGGAAGGAGACAGACGTCATCCCAATTTCTTGATGACAGATCGA 316
QY 111 ArgAsnSerSerGluGluArgSerArgProAsnSerGlyAspGluLysArgLysLeuSer 130
Db 315 CGTAATTCAGTGAGGAGAGAGAGTAGGCCAACTCAGTGTATGAGAGCGCAAACTTAGC 256
QY 131 LeuSerGluGlyLysValLysLysAsnMetGlyGlyGlyLysArgValSerSerGly 150
Db 255 CTGTCTGAAGGGAAGGTGAAGAGAACATGGGAGAGGAAGAGAGGGGTGTCTTCAGGC 196
QY 151 SerSerGluArgGlySerLysSerProLeuLysArgAlaGlnGlnSerProArgLys 170
Db 195 TCTTCAGAGAGAGCTCCAATCCCTCTGAAGAGAGCCCAAGAGCAAGTCCCGGAAG 136
QY 171 ArgGlyArgProLysAspGluLysAspLeuThrIleProGluSerSerThrValLys 190

```

```

Db 135 CGGGGTCCGCCCCCAAGGATGAGAGGATCTCACCATCCCGAGTCTAGTACCGTGAAG 76
QY 191 GlyMetMetAlaGlyProMetAlaAlaPheLysTrpGlnProThrAla---SerGluPro 209
Db 75 GGGATGATGCCCGGACCGATGCGCGGTTTAATGGCAGCAACCGGCTCTCTAGGCCA 16
QY 210 ValLys 211
Db 15 TGAAG 10

RESULT 10
AC020663/c
LOCUS AC020663 154218 bp DNA linear PRI 08-JAN-2000
DEFINITION Homo sapiens chromosome 16 clone RPCI-11_127120, complete sequence.
ACCESSION AC020663
VERSION AC020663.1 GI:6682593
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 154218)
AUTHORS Bruce, D., Mundt, M., Doggett, N., Munk, C., Saunders, E., Robinson, D.,
Jones, M., Buckingham, J., Chasteen, L., Thompson, S., Goodwin, L.,
Bryant, J., Tesmer, J., Meincke, L., Longmire, J., White, S., Tatum, O.,
Campbell, C., Fawcett, J., Maltbie, M., Bussod, M., Sutherland, R.,
Han, C. and Deaven, L.
TITLE Sequencing of Human Chromosome 16q12
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 154218)
AUTHORS Ricke, D.O.
TITLE Large Scale Sequence Analysis and Annotation with the Sequence
Comparison Analysis (SCAN) System
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 154218)
AUTHORS Bruce, D., Mundt, M., Doggett, N., Munk, C., Saunders, E., Robinson, D.,
Jones, M., Buckingham, J., Chasteen, L., Thompson, S., Goodwin, L.,
Bryant, J., Tesmer, J., Meincke, L., Longmire, J., White, S., Tatum, O.,
Campbell, C., Fawcett, J., Maltbie, M., Bussod, M., Sutherland, R.,
McMurry, K., Han, C. and Deaven, L.
TITLE Direct Submission
JOURNAL Submitted (08-JAN-2000) Center for Human Genome Studies, DOE Joint
Genome Institute, Los Alamos National Laboratory, MS M888, Los
Alamos, NM 87545, USA
FEATURES
source
1..154218
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosomes="16"
/clone="RPCI-11_127120"
/complement(41..330)
repeat_region /rpt_family="Alu"
repeat_region 575..882
repeat_region /rpt_family="Alu"
repeat_region 1414..1712
repeat_region /rpt_family="Alu"
repeat_region 1595..1862
repeat_region /rpt_family="Alu"
repeat_region 1871..2190
repeat_region /rpt_family="Alu"
repeat_region 2176..2372
repeat_region /rpt_family="L1"
repeat_region 2498..2795
repeat_region /rpt_family="Alu"
repeat_region 3187..3504
repeat_region /rpt_family="Alu"
repeat_region complement(3571..3890)
repeat_region /rpt_family="Alu"
repeat_region complement(3859..4163)
repeat_region /rpt_family="Alu"
repeat_region 4198..4503

```

```
repeat_region /rpt_family="Alu" 25205. .25461
repeat_region /rpt_family="Alu" 25605. .25862
repeat_region /rpt_family="L1" 25914. .25981
misc_feature complement(6848. .6983)
repeat_region /note="GRAIL 2 excellent exon, frame 1" 26335. .26355
repeat_region /rpt_family="Alu" 28437. .28732
repeat_region /rpt_family="Alu" 30597. .30772
misc_feature /note="GRAIL 2 excellent exon, frame 2" 31300. .31655
repeat_region /rpt_family="Alu" 32059. .32124
repeat_region /rpt_family="Alu" 32492. .32492
repeat_region /rpt_family="Alu" 32507. .32839
repeat_region /rpt_family="Alu" 33649. .34040
repeat_region /rpt_family="Alu" 34279. .34556
repeat_region /rpt_family="LTR37B" 35017. .35495
repeat_region /rpt_family="Alu" 36991. .36991
misc_feature /note="GRAIL 2 excellent exon, frame 0" 37667. .37808
repeat_region complement(37920. .37961)
repeat_region /rpt_family="MIR" 38430. .38745
repeat_region /rpt_family="Alu" 38789. .38899
repeat_region /rpt_family="MER3" 40939. .41089
repeat_region /rpt_family="Alu" 42294. .42589
repeat_region /rpt_family="Alu" 4405. .44482
repeat_region /rpt_family="MLTIC" 44515. .44693
repeat_region /rpt_family="Alu" 45916. .46254
repeat_region /rpt_family="Alu" 46347. .46347
misc_feature /note="GRAIL 2 excellent exon, frame 1" 46426. .46596
repeat_region /rpt_family="MER20" 47139. .47448
repeat_region /rpt_family="Alu" 47535. .47943
repeat_region /rpt_family="THE1" 48377. .48377
repeat_region /rpt_family="Alu" 48431. .48778
repeat_region /rpt_family="L1" 48877. .49260
repeat_region /rpt_family="Alu" 49310. .49588
repeat_region complement(49310. .49588)

Alignment Scores:
Pred. No.: 8.72e-28 Length: 154218
Score: 551.00 Matches: 128
Percent Similarity: 40.92% Conservative: 5
Best Local Similarity: 39.38% Mismatches: 8
```

```
Query Match: 41.77% Indels: 184
DB: 9 Gaps: 1
US-09-987-755-2 (1-249) x AC020663 (1-154218)

Qy 53 AlaTrrpIleLysValGluGlnLeuLysProTyrHisAlaHisLysGluGluMetIleLys 72
Db 125672 GCCTGGATCAAGTGAACAGCTGAAGCCATATCATGCTCATAAAGAGAAATGATAAAA 125613

Qy 73 IleAsnLysGlyLysArgPheGlnGlnAlaValAlaValGluGluPheLeuArgG 92
Db 125612 ATTAACAGGGTAAACATTCACAGACGGTAGATGCTGTGGAAGAGTTCTCCAGAGA 125553

Qy 93 AlalysGlyLysAsp----- 97
Db 125552 GCCAAGGGAAGACACAGGTGAGAGACACAATGTTCACTTCTTACCAATGCCAGACTGT 125493

Qy 97 ----- 97
Db 125492 GATTTTGCATAGCACTAACAGATTAAATGTCACCAACGCAACCTTTGTTCTCTG 125433

Qy 97 ----- 97
Db 125432 AAATGCTACTGGAAAGCTACTTCACTTAAGGTGTACGGTTTGTCTTATCAGTTG 125373

Qy 97 ----- 97
Db 125372 AAATCCATGCAGTAACCTTCTGGAGTAGTTGTTGCTTTTAGGAAATTCGTTGGAC 125313

Qy 97 ----- 97
Db 125312 CCAGGAGAAATATCTCAGCATTTGAGTCTAGCACAGGCCATTTCTCTTTCATAGACA 125253

Qy 97 ----- 97
Db 125252 CAGATTAGAGCATAGCCTCTGAATAACATATCTGCTAGATTAAATCCTTTCTTCTA 125193

Qy 97 ----- 97
Db 125192 GTTTTGGACATAGGAAAGTGACTTTTGATATATATTTTAAATCTGCTTCTCCA 125133

Qy 97 ----- 97
Db 125132 ATGCTGGACGACAGCAGCAATCTATTCTAGGTCATTTGTAGGAGACACACAGAGTG 125073

Qy 97 ----- 97
Db 125072 CCTGTAGCAGCATGGCACTGTAAATGATGGTGGCGATGGCGGACGCTGTGTGG 125013

Qy 98 -----GlnThrSerSerHisAsnSerSerAspLys 108
Db 125012 TAGTTATCATTATTATTATTAACTTTTCAGACGTCATCCCAATTTCTTGATGACAG 124953

Qy 109 AsnArgArgAsnSerSerGluGluArgSerArgProAsnSerGlyAspGluLysArgLys 128
Db 124952 AATCGAGTAAATTCAGTGGAGAGAGTAGGCGCAAACTCAGGTGATGAGAGCGCAAA 124893

Qy 129 LeuSerLeuSerGluGlyLysValLysLysAsnMetGlyGluGlyLysArgValSer 148
Db 124892 CTTAGCCTGTCTGAAGGGAAGGTGAAGAACAATGGGAGAGGCAAGAGAGGTGTCT 124833

Qy 149 SerGlySerSerGluGlyLysSerLysSerProLeuLysArgAlaGlnGluGlnSerPro 168
Db 124832 TCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTGAAAGAGCCCAAGCAAGTCCC 124773

Qy 169 ArgLysArgGlyArgProProLysAspGluLysAspLeuThrIleProGluSerSerThr 188
Db 124772 CGGAAGCGGGTCCGCCCCCAAGAGATGAGAAGGTTGTTTCTCTCTGCAGCTTCCGGG 124713

Qy 189 ValLysGlyMetMet 193
Db 124712 ATTAGTGATGTCCTT 124698
```

```
RESULT 11
AC027687/c
LOCUS
DEFINITION Homo sapiens chromosome 16 clone RP11-61L4, complete sequence.
AC027687
AC027687.7 GI:29294025
HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 169166)
DOE Joint Genome Institute, Stanford Human Genome Center and Los
Alamos National Laboratory.
Direct Submission
Unpublished
2 (bases 1 to 169166)
DOE Joint Genome Institute.
Direct Submission
Submitted (01-APR-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 169166)
DOE Joint Genome Institute.
Direct Submission
Submitted (01-MAR-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
4 (bases 1 to 169166)
DOE Joint Genome Institute, Stanford Human Genome Center and Los
Alamos National Laboratory.
Direct Submission
Submitted (27-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Mar 27, 2003 this sequence version replaced gi:19033441.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center and Los Alamos
National Laboratory
www.shgc.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence;
Estimated Total Number of Errors is 0.
FEATURES
source
1. 169166
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-61L4"
BASE COUNT 41124 a 42936 c 42004 g 43102 t
ORIGIN
Alignment Scores:
Pred. No.: 9.66e-28 Length: 169166
Score: 551.00 Matches: 128
Percent Similarity: 40.92% Conservative: 5
Best Local Similarity: 39.38% Mismatches: 8
Query Match: 41.77% Indels: 184
DB: 9 Gaps: 1
US-09-987-755-2 (1-249) x AC027687 (1-169166)
Qy 53 AlaTrrpIleLysValGluGlnLeuLysProTyrHisAlaHisLysGluGluMetIleLys 72
Db 25492 GCCTGGATCAAGTGAACAGCTGAAGCCATATCATGCTCATAAAGAGAAATGATAAAA 25433
Qy 73 IleAsnLysGlyLysArgPheGlnGlnAlaValAlaValGluGluPheLeuArgG 92
Db 25432 ATTAACAGGGTAAACATTCACAGACGGTAGATGCTGTGGAAGAGTTCTCCAGAGA 25373
Qy 93 AlalysGlyLysAsp----- 97
Db 25372 GCCAAGGGAAGACACAGGTGAGAGACACAATGTTCACTTCTTACCAATGCCAGACTGT 25313
Qy 97 ----- 97
```

Db 25312 GATTTTGCATAGCAACTAACAGATTAAATGTACCCAAACGAAACCTTTTCTTCTG 25253
QY 97 ----- 97
Db 25252 AAATCTACTGGAGAAAGCTACTCTCAGTCTTAAGTGTACGGTTTTCCTTATCAGTTTG 25193
QY 97 ----- 97
Db 25192 AAATCCCATGCAGCTAACCTTTCTGGAGTAGTGTGTCTTTTAGGAAAAATTGCTTGGAC 25133
QY 97 ----- 97
Db 25132 CCAGGAGAAATATCTCAGCATTTGAGTCTAGCACAGGCCATTCTTCTTCATAGCACACA 25073
QY 97 ----- 97
Db 25072 CAGATTAGGAGCATGACCTCTGAAATACATATATCTAGTCTGCTAGATTAAATCTTCTTCTTA 25013
QY 97 ----- 97
Db 25012 GTTTTGGGACATAGGAAAGTGACCTTTTGATATATATATTTTAAATCTGCTTCTCCCA 24953
QY 97 ----- 97
Db 24952 ATGCTGGGACGACAGCAATCTATTCTAGGTCATTGTGAGGAGACCCACACAGAGTG 24893
QY 97 ----- 97
Db 24892 CCTGTAGCAGCATGGCACCTGTAAATGATGTGGCGATGGCGGCACGACTGTGGTG 24833
QY 98 -----GlnThrSerSerHisAenSerSerAspAspLys 108
Db 24832 TAGTTATCATTTATTATTAACTTTTCAGAGCTCATCCCAATTTCTTGATGACAAAG 24773
QY 109 AenArgArgAenSerSerGluArgSerArgProAenSerGlyAspGluLysArgLys 128
Db 24772 AATGACGCTAAATTCAGTGAGGAGAGAGTAGTGGCCAACTCAGGTGATGAGAGGCGCAA 24713
QY 129 LeuSerLeuSerGluGlyLysValLysValLysValLysValLysValLysValLys 148
Db 24712 CTTAGCTCTGTAAGGAGAGGTGAAGAGAAACATGGAGAGAGGAAAGAGGGTGTCT 24653
QY 149 SerGlySerSerGluArgGlySerLysSerProLeuLysArgAlaGlnGlnSerPro 168
Db 24652 TCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTGAAAGAGGCCAAGCAAGTCCC 24593
QY 169 ArgLysArgGlyArgProProLysAspGluLysAspLeuThrIleProGluSerSerThr 188
Db 24592 CGGAAGCGGGGTGGGCCCCCAAGGATGAGAAGGTTTGTCTCTCTGCAGCTTCCGG 24533
QY 189 ValLysGlyMetMet 193
Db 24532 ATTAGTGATGCTT 24518

RESULT 12
LOCUS AC113755/c
DEFINITION Rattus norvegicus clone CH230-193Cl, *** SEQUENCING IN PROGRESS
ACCESSION AC113755
VERSION AC113755.5 GI:25072692
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 233786)
AUTHORS Murzyn D, Marie, Metzker M, Lee, Abramson S, Adams C, Alder J, Allen C, Allen H, Albrooks S, Amin A, Anguiano D, Anyalebechi V, Aoyagi A, Ayodeji M, Baca E, Baden H, Baldwin D, Bandaranaike D, Barber M, Barnstead M, Benahmed F,

Biswalo K, Blair J, Blankenburg K, Blyth P, Brown M, Bryant N, Buhay C, Burch P, Burrell K, Calderon E, Cardenas V, Carter K, Cavazos I, Ceasar H, Center A, Chacko J, Chavez D, Chen G, Chen Y, Chen Z, Chu J, Cleveland C, Cockrell R, Cox C, Coyle M, Cree A, D'Souza L, Davila M, Davis C, Davy-Carroll L, De Anda C, Dederich D, Delgado O, Denson S, Deramo C, Ding Y, Din H, Divya K, Egan A, Escotto M, Eugene C, Evans C, Falls T, Fan G, Fernandez S, Finley M, Flagg N, Forbes L, Foster M, Foster P, Fraser C, Gabisi A, Ganta R, Garcia A, Garner T, Garza M, Gregeorgis E, Geir K, Gill R, Grady M, Guerra W, Gunaratne P, Haaland W, Hamil C, Hamilton C, Hamilton K, Harvey Y, Havlak P, Hawes A, Henderson N, Hernandez J, Hernandez R, Hines S, Hladun S, Hodgson A, Hogues M, Hollins B, Howells S, Hulyk S, Hume J, Idlebird D, Jackson A, Jackson L, Jacob L, Jiang H, Johnson B, Johnson R, Jolivet A, Karpachy S, Kelly S, Kelly S, Khan Z, King L, Kovar C, Kowis C, Kraft C, Lebow H, Levan J, Lewis L, Li Z, Liu J, Liu J, Liu W, Liu Y, London P, Longacre S, Lopez J, Lorensuhewa L, Loulseged H, Lozado R, Lu X, Ma J, Maheshwari M, Mahindartne M, Mahmoud M, Malloy K, Mangum A, Mangum B, Mapua P, Martin K, Martin R, Martinez E, Mawhiney S, McLeod M, McNeil T, Meenen E, Milosavljevic A, Miner G, Minja E, Montemayor J, Moore S, Morgan M, Morris K, Morris S, Munidasa M, Murphy M, Nair L, Nankervis C, Neal D, Newton N, Nguyen N, Norris S, Nwaokemele O, Okwuonu G, Olarnpunsagoon A, Pal S, Parke K, Pasternak S, Paul H, Perez A, Perez L, Pfannkoch C, Popper F, Poindexter A, Popovic D, Primus E, Pu L, Puazo M, Quiroz J, Rachlin E, Reeves K, Regier M, Reigh R, Reilly B, Reilly M, Ren Y, Reuter M, Richards S, Riggs F, Rives C, Rodkey T, Rojas A, Rose M, Rose R, Ruiz S, Sanders W, Savery G, Scherer S, Scott G, Shatsman S, Shen H, Shetty J, Shvartsbeyn A, Sisson I, Sitter C, Smajs D, Sneed A, Sodergren E, Song X, Sorelle R, Sosa J, Steimle M, Strong R, Sutton A, Svatek A, Tabor P, Taylor C, Taylor T, Thomas N, Thomas S, Tingey A, Trejos Z, Usmani K, Valas R, Vera V, Villalana D, Waldron L, Walker B, Wang J, Wang Q, Wang S, Warren J, Warren R, Wei X, White F, Williams G, Willson R, Wleczek R, Wooden H, Worley K, Wright D, Wright R, Wu J, Yakub S, Yen J, Yoon L, Yoon V, Yu F, Zhang J, Zhou J, Zhou X, Zhao S, Dunn D, von Niederhausern A, Weiss R, Smith D, Holt R, Smith R, Smith H, O., Weinstock G, and Gibbs R.A.
Direct Submission
Unpublished
2 (bases 1 to 233786)
Worley K.C.
Direct Submission
Submitted (05-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 233786)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Nov 19, 2002 this sequence version replaced gi:23267215.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
----- Genome Center

Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GSEQ
 Center clone name: CH230-193C1

----- Summary Statistics

Assembly program: Phrap; version 0.990329
 Consensus quality: 216369 bases at least Q40
 Consensus quality: 218310 bases at least Q30
 Consensus quality: 219420 bases at least Q20
 Estimated insert size: 221075; sum-of-contigs estimation
 Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 3 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 11093: contig of 11093 bp in length
 * 11094 11193: gap of unknown length
 * 11194 232615: contig of 221422 bp in length
 * 232616 232715: gap of unknown length
 * 232716 233786: contig of 1071 bp in length.

FEATURES

source

1. 233786
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-193C1"

misc_feature

1. 11153
 /note="wgs end extension"

misc_feature

978..1446
 /note="clone boundary"

clone end:T7

site:

end sequence:BH309210"

8020..11093

/note="wgs contig"

229660..230825

/note="wgs contig"

231327..232615

/note="wgs contig"

BASE COUNT 55548 a 52773 c 53568 g 58339 t 13558 others

ORIGIN

Alignment Scores:

Pred. No.: 5.46e-25 Length: 233786
 Score: 514.00 Matches: 123
 Percent Similarity: 44.56% Conservative: 7
 Best Local Similarity: 43.16% Mismatches: 151
 Query Match: 38.97% Indels: 151
 DB: 2 Gaps: 3

US-09-987-755-2 (1-249) x AC113755 (1-233786)

Qy 39 LysLysCysPheValLysPheGlyThrGluAspHisAlaTrpIleLysValGlu 58

Db 204143 AAACCTGTTTGT-----CCTACTAGCGCATGGATCAAAAGTGGAG 204102

Qy 59 GlnLeuLysProTyrHisAlaHisLysGluMetIleLysIleAsnLysGlyLysArg 78

Db 204101 CAGCTGAAGCCTTACCATGCTCAAGAGAGAGATGATCAAGATCAACAGGGCAGCGG 204042

Qy 79 PheGlnGlnAlaValAspAlaValGluPheLeuArgArgAlaLysGlyLysAsp--- 97

Db 204041 TTCCAGCAAGCTGTGGATGCTGTGAGGAGATTCTCTCAGGAGAGCCAAAGGCAAGACCAG 203982

Qy 97 ----- 97

Db 203981 GTGAGAGAGAGAGGGCTGCAGTGGTCCCTCCAACATCAGGCCAGTGGTTTGCATGTTTC 203922

Qy 97 ----- 97

Db 203921 TGAGAAGATGTTTGTGTCTAGACACTCTTTTCCGTGAAATGCTTAGAAGAGGACACCTG 203862

Qy 97 ----- 97

Db 203861 TGTCCAGAACTCTGCCCGTGGCTCCCTTTTCTAAGCGATGTTGTTTGGGAGAGATCGC 203802

Qy 97 ----- 97

Db 203801 TCACCTTTCACAGCACAGGCAAGAGCTTAGAGTACAGCCACCTCTCACCTTTGGGGCTG 203742

Qy 97 ----- 97

Db 203741 TAGACAAGCGCATGTTTCTTTTAAAGCTGTCTTTTCCAGCACAAAGAGAGGTGAGC 203682

Qy 97 ----- 97

Db 203681 AAGGATCTGTTTCTCAGGGTGTATGAAGACAATTACAGAACTCTGAGAGAACTGTG 203622

Qy 97 ----- 97

Db 203621 TGGCCCATGTTGATGCGCAGTGGTATGATGCTGTTGTAGTGTATCATTTATTGTT 203562

Qy 98 -----GlnThrSerSerHisAsnSerSerAspLysAsnArgAsnSerSerG1 115

Db 203561 GAACCTTTCAGACATCATCTCACACTTCTGCTGATGACAAGATCGCGTAAATCCAGTGA 203502

Qy 115 uGluArgSerArgProAsnSerGlyAspGluLysArgLysLeuSerLeuSerGluGly 135

Db 203501 GGAGAGAAGTAGGCCAAACTCAGGTGATGAGAAACGCAAGCTTAGCTGTCTGAAGGAA 203442

Qy 135 sVallysLysAsnMetGlyGluLysLysArgValSerSerGlySerSerGluArgG1 155

Db 203441 GGTGAAGAGAACATCGGAGAGAAAGAGAGGTGACTTCAGGCTCTCAGACAGAG 203382

Qy 155 YSerLysSerProLeuLysArgAlaGlnGluInSerProArgLysArgGlyArgProPr 175

Db 203381 CTCCAAATGC---CTTAAAGAGCCCAAGAGCAAGTCCCGGAAGCGGGTGGCCCCC 203325

Qy 175 oLysAspGluLys 179

Db 203324 AAAGGATGAGAAG 203312

RESULT 13

AC108242 239081 bp DNA linear HTG 13-NOV-2002

LOCUS

DEFINITION Rattus norvegicus clone CH230-230G18, *** SEQUENCING IN PROGRESS

***, 2 unordered pieces.

ACCESSION AC108242

VERSION AC108242.5 GI:24942514

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 239081)

AUTHORS Muzny, D., Maric, Metzkner, M., Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Chen, H., Chao, J., Chacko, J., Chavez, D., Chen, G., Chen, Y., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,

Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escoto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hodges, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuewa, L., Loulseghe, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhiney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K., Nwaoketeme, O., Okwodu, G., Olampunsgoon, A., Pal, S., Pankoch, C., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankoch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, P., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steidle, M., Strong, R., Sutton, A., Svatek, R., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Direct Submission
Unpublished

2 (bases 1 to 239081)

Worley, K.C.

Submitted (27-JAN-2002)

Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 239081)

Rat Genome Sequencing Consortium.

Submitted (13-NOV-2002)

Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 13, 2002 this sequence version replaced gi:23119074.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

Center: Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: GJRI
Center clone name: CH230-230G18
----- Summary Statistics
Assembly program: Phrap: version 0.990329
Consensus quality: 225503 bases at least Q40
Consensus quality: 228480 bases at least Q30
Consensus quality: 230425 bases at least Q20
Estimated insert size: 233061; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 5876: contig of 5876 bp in length
* 5877 5976: gap of unknown length
* 5977 239081: contig of 233105 bp in length.

FEATURES
source

1. 239081

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-230G18"

1. 2515

/note="wgs end extension"

clone_end:T7"

4849..5674

/note="clone boundary"

clone_end:T7

site:

end sequence:RWBOF45TJB"

5977..7403

/note="wgs contig"

complement(230992..231789)

/note="clone boundary"

clone_end:Sp6

site:

end sequence:RWBOF45TVB"

61633 a 56132 c 55001 g 58761 t 7554 others

BASE COUNT

ORIGIN

Alignment Scores:

Pred. No.: 5,6e-25 Length: 239081

Score: 514.00 Matches: 123

Percent Similarity: 44.56% Conservative: 4

Best Local Similarity: 43.16% Mismatches: 7

Query Match: 38.97% Indels: 151

DB: 2 Gaps: 3

US-09-987-755-2 (1-249) x AC108242 (1-239081)

Qy 39 LysLysCysPhePheValLysPhePheGlyThrGluAspHisAlaTrpIleLysValGlu 58
Db 93353 AAACCTGTTTGTGTTT-----CCTACTAGCGCATGATCAAAAGTGGAG 93394

Qy 59 GlnLeuLysProTyrHisAlaHisLysGluMetIleLysIleAsnLysGlyLysArg 78
Db 93395 CAGCTGAAGCTTACCATGCTCACAGAGAGAGATGATCAAGATCAACAGGGCAAGCGG 93454

Qy 79 PheGlnGlnAlaValAspAlaValGluGluPheLeuArgAlaLysGlyLysAsp--- 97
Db 93455 TTCCAGCAAGCTGTGGATGCTGTGCGAGGAGTTCTCTCAGGAGAGCCAGGCAAGACCAG 93514

Qy 97 ----- 97
Db 93515 GTGAGAGAGAGAGGGGCTGAGTGTGCTCCACATCAGGCCAGTGGTTTGTGATGTTTC 93574


```

Qy 97 ----- 97
Db 93575 TGAGAAGATGTTGTGTGTCAGACACTCTTTTCGTCGTAATGCTCCTAGAGAGACACCTG 93634
Qy 97 ----- 97
Db 93635 TGTCAGAACTCTGCCGTGGCTCCCTTTTCTAAGGATGTTGTTTGGAGAGATGCG 93694
Qy 97 ----- 97
Db 93695 TCACCTTCACAGCACAGCGAAGAGCTAGGAGTACAGCACCTCTCACCTTTGGGCTG 93754
Qy 97 ----- 97
Db 93755 TAGACAAGCGACATGTTTCTTTTAAAGTCTCTTTCCAGCACAAAGAGAGGTCAGC 93814
Qy 97 ----- 97
Db 93815 AAGGATCCTGTTTCTCAGGGTGTATGAACAATTCACAGAACTGCTGAGAACTGTG 93874
Qy 97 ----- 97
Db 93875 TGGCCCATGTTGATGGCAGTGGTATAGTGGCTGTGTAGTGGTTATCATTTATTGTT 93934
Qy 98 -----GlnThrSerSerHisAsnSerSerAspAspIysAsnArgAsnSerSerG1 115
Db 93935 GAACTTTTCAGACATCATCTCACACTTCTGCTGATGACAAAGAAATCGGCTAATTCAGTGA 93994
Qy 115 uGluArgSerArgProAsnSerGlyAspGluLysArgLysLeuSerLeuSerGluGly 135
Db 93995 GGAGAGAAGTAGGCGCAACTCAGGTGATGAGAAACGCAAGCTTACCTGCTGTAAGGGA 94054
Qy 135 sValLysLysAsnMetGlyGluGlyLysLysArgValSerSerGlySerSerGluArgG1 155
Db 94055 GGTCAAGAAAGAAATGGGAGAGAGAAAGAGAGGGTACCTTCAGGCTCTGACAGACAGG 94114
Qy 155 ySerLysSerProLeuLysArgAlaGlnGlnSerProArgLysArgGlyArgProPr 175
Db 94115 CTCCAAATGC---CTTAAAGAGAGCCCAAGAGCAAGTCCCGGAAGCGGGTGGCCCCC 94171
Qy 175 OlyAspGluLys 179
Db 94172 AAAGGATGAGAAG 94184

RESULT 14
LOCUS AC123492 243770 bp DNA linear HTG 12-OCT-2002
DEFINITION Rattus norvegicus clone CH230-100L13, WORKING DRAFT SEQUENCE.
ACCESSION AC123492
VERSION AC123492.3 GI:23666397
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
AUTHORS 1 (bases 1 to 243770)
Murny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Albrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Dexamco,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,

```

Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpthy, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.B., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhewa, L., Loulseged, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Hawthiney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Naif, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokemele, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartabeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhauser, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstein, G. and Gibbs, R.A.

Direct Submission
 Unpublished
 2 (bases 1 to 243770)
 Worley, K.C.
 Direct Submission
 Submitted (29-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 243770)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (12-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On Oct 10, 2002 this sequence version replaced gi:21908419.
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
 Center: Baylor College of Medicine
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GXR
 Center clone name: CH230-100L13
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990129
 Consensus quality: 227678 bases at least Q40
 Consensus quality: 229865 bases at least Q30
 Consensus quality: 231324 bases at least Q20

Estimated insert size: 237575; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

- * NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
- * NOTE: This is a 'working draft' sequence. It currently
consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces
is believed to be correct as given, however the sizes
of the gaps between them are based on estimates that have
been provided by the submitter.
- * This sequence will be replaced
by the finished sequence as soon as it is available and
the accession number will be preserved.
- * 1 243770: contig of 243770 bp in length.

FEATURES

source

1. 243770
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-100L13"

misc_feature

1
/note="clone boundary
site: EcoRI
end: 17"

misc_feature

complement(240521..241363)
/note="clone boundary
clone end: Sp6
site: EcoRI
end: sequence: BH305306"

BASE COUNT 60464 a 56544 c 55414 g 59687 t 11661 others
ORIGIN

Alignment Scores:

Pred. No.: 5,728-25 Length: 243770
Score: 514.00 Matches: 123
Percent Similarity: 44.56% Conservative: 4
Best Local Similarity: 43.16% Mismatches: 7
Query Match: 38.97% Indels: 151
DB: 2 Gaps: 3

US-09-987-755-2 (1-249) x AC123492 (1-243770)

Qy 39 LysLysCysPheValLysPheGlyThrGluAspHisAlaTrpIleLysValGlu 58
Db 150572 AAACCTGTTCTTT-----CCTACTAGCGCATGATCAAGTGGAG 150613
Qy 59 GlnLeuLysProTyHisAlaHisLysGluMetIleLysIleAsnLysGlyLysArg 78
Db 150614 CAGCTGAAGCCTTACCATGCTCACAAAGGAGGATGATCAAGATCAACAAAGGCAAGCGG 150673
Qy 79 PheGlnGlnAlaValAspAlaValGluPheLeuArgAlaLysGlyLysAsp--- 97
Db 150674 TTCCAGCAAGCTGTGGTGTCTGTGAGAGTTCTTCAGGAGGCAAGGCAAGACCAG 150733
Qy 97 ----- 97
Db 150734 GTGAGAGAGAGGGGCTGCAGTGTCTCCACATCAGGCCAGTGGTTGCATGGTTC 150793
Qy 97 ----- 97
Db 150794 TGAGAAGATGGTTGTGTGCAGACACTCTTTTCGTGAATGCCTAGAAAGAGACACCTG 150853
Qy 97 ----- 97
Db 150854 TGTCCAGAACTCTGCCCGTGGCTCCCTTTTCTAAGGCATGTTGTTTGGGAGGATGCG 150913
Qy 97 ----- 97
Db 150914 TCACCTTCACAGCACAGGCAAGAGCTTAGGAGTACAGCCACCTCTACCTTTGGGGCTG 150973
Qy 97 ----- 97

Db 150974 TAGACAAGCGACATGTTTCTTTTAAAGCTGTCTTTCCAGCAACAAGAGAGGTGAGC 151033
Qy 97 ----- 97
Db 151034 AAGGATCCTGTTTCTCAGGGTGCTATGAGACAATTCACAGAACTGCTGAGAGAACTGTG 151093
Qy 97 ----- 97
Db 151094 TGGCCCATGTTGATGGCAGTGGTATAGTGCTGTTGTAGTGGTTATCATTTATTATGTT 151153
Qy 98 -----GlnThrSerSerHisAsnSerSerAspLysAsnArgAsnSerSerG1 115
Db 151154 GAACCTTTTCAGACATCATCTCAGCTCTGCTGATGACAAGAAATCGCGTAATTCAGTGA 151213
Qy 115 uGluArgSerArgProAsnSerGlyAspGluLysArgLysLeuSerLeuSerGluGlyL 135
Db 151214 GGAGAGAGTAGGCAAACTCAGGTGATGAGAAACGCAAGCTTAGCTGTCTGAAGGAA 151273
Qy 135 sValLysAsnMetGlyGluGlyLysLysLysValSerSerGlySerSerGluArgG1 155
Db 151274 GGTGAAGAAGAAATCGGAGAGGAAAGAGAGGTGACTTCAGGCTCTGCAGACAGAGG 151333
Qy 155 ySerLysSerProLeuLysArgAlaGlnGluSerProArgLysArgGlyArgPropR 175
Db 151334 CTCCAATGTC--CTTAAGAGAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTCGGCCCC 151390
Qy 175 oLysAspGluLys 179
Db 151391 AAAGATGAGAG 151403

RESULT 15

AC127246/c

LOCUS

AC127246

DEFINITION

Mus musculus chromosome 16 clone RP24-424L20, complete sequence.

ACCESSION

AC127246.3

VERSION

GI:28626893

KEYWORDS

HTG.

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

McPherson, J.D. and Waterston, R.H.

The sequence of Mus musculus clone

Unpublished

2 (bases 1 to 181475)

McPherson, J.D. and Waterston, R.H.

Direct Submission

Submitted (14-JUL-2002) Genome Sequencing Center, 4444 Forest Park

Parkway, St. Louis, MO 63108, USA

3 (bases 1 to 181475)

McPherson, J.D. and Waterston, R.H.

Direct Submission

Submitted (29-JAN-2003) Genome Sequencing Center, 4444 Forest Park

Parkway, St. Louis, MO 63108, USA

4 (bases 1 to 181475)

McPherson, J.D. and Waterston, R.H.

Direct Submission

Submitted (01-MAR-2003) Genome Sequencing Center, 4444 Forest Park

Parkway, St. Louis, MO 63108, USA

On Mar 1, 2003 this sequence version replaced gi:28015415.

----- Genome Center -----

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc/index.shtml>Contact: submissions@wustl.edu

----- Project Information -----

Center project name: M.BB0424L20

----- Location/Qualifiers -----

1. 181475

FEATURES

source

/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="16"
/clone="RP24-424L20"
BASE COUNT 46692 a 42460 c 43716 g 48607 t
ORIGIN

Alignment Scores:
Pred. No.: 1.64e-24 Length: 181475
Score: 505.50 Matches: 126
Percent Similarity: 44.41% Conservatives: 5
Best Local Similarity: 42.71% Mismatches: 12
Query Match: 38.32% Indels: 153
DB: 10 Gaps: 4

US-09-987-755-2 (1-249) x AC127246 (1-181475)

```
Qy 43 PheValLysPheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeuLysPro 62
Db 38693 TTTCTGTGTTTGTCTTCTACTAGT-----GCCTGGATCAAAGTGAACAGCTAAAGCCT 38640

Qy 63 TyrHisAlaHisLysGluMetIleLysIleAsnLysGlyLysArgPheGlnAla 82
Db 38639 TACCATGCTCACAAGGAGGAGATATAAGATTAAACAAAGGTAAACGGTTCCAGCAAGCT 38580

Qy 83 ValAspAlaValGluGluPheLeuArgAlaLysGlyLysAspGln----- 98
Db 38579 GTGGATGCTGTTGAAGAGTTCCTCAGGAGGCCAAAGGAAAGACCA-GGTGAGAGAGGC 38521

Qy 98 ----- 98
Db 38520 TGCAGTGTTCCTCCCAACATCGGCTCAATGATTTTTCATGTTTACTAGCAGATGGTG 38461

Qy 98 ----- 98
Db 38460 GTGTCAATAGACAGCTCCTTTCCCATGAGATTTCCCTAGCAAGAAACAGCTGTGTCCAG 38401

Qy 98 ----- 98
Db 38400 AACTATATGCTGTGGCTCACTTTTCTGAGTCATGCATGTTGTTTCGGAAGAGTAGCCC 38341

Qy 98 ----- 98
Db 38340 AGCTTCACATACAGGAAGAGATCTTAGGAATTCAGCCACCTTTGGGGCTGTAGACAACT 38281

Qy 98 ----- 98
Db 38280 GAGGTGATTTTTTTTTTTTAAAGTGTCTTTCCAGCACAAACAGAGGTGAGCAAGGATG 38221

Qy 98 ----- 98
Db 38220 CTATTTCTCAGGGTGGCTATGAAGACAGTTTCACAGAACTGCTGAGAGAGAGAGAACTG 38161

Qy 98 ----- 98
Db 38160 TGTGCCCATGTTGATGCGCAGTGTGTAGTGGCTGTTGTAGTGTATCATTATTATTG 38101

Qy 99 -----ThrSerSerHisAsnSerSerSerAspLysAsnArgArgAsnSerSer 114
Db 38100 TTGAACCTTTTCAGACATCATCCACACTTCTGCTGATGACAAAGATCGGCGTAATTCAGT 38041

Qy 115 GluGluArgSerArgProAsnSerGlyAspGlyLysArgLysLeuSerLeuSerGluGly 134
Db 38040 GAGGAGAGAGTAGGCCAACTCAGGTGATGAGAAACCGCAAGCTTAGCCTGTCTGAAGGG 37981

Qy 135 LysValLysLysAsnMetGlyGluGlyLysLysArgValSerSerGlySerSerGluArg 154
Db 37980 AAGGTGAAGAAGAACATGGGAGAGAGGAAGAGAGGGTGACTTCAGGCTCTGCACACAGA 37921

Qy 155 GlySerLysSerProLeuLysArgAlaGlnGluGlnSerProArgLysArgGlyArgPro 174
Db 37920 GGCTCCAAATGC---CTTAAAGAGCCCAAGAGCAAGTCCCGGAAAGCGGGTCCGCC 37864
```

Qy 175 ProLysAspGluLys---AspLeuThrIleProGluSerSerThr 188
Db 37863 CCAAAGGATGAGAAGGTTTGTCTTACTGCGGTAGCTTCTGGGACT 37819

Search completed: December 8, 2003, 12:46:49
Job time : 3978 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 8, 2003, 10:41:11 ; Search time 325 Seconds
(without alignments)

2546.388 Million cell updates/sec

Title: US-09-987-755-2

Perfect score: 1319

Sequence: 1 MAASVRLGLVWGKLGYP.....VCYQAITKKLKICEDLLPR 249

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Xgapop 10.0 , Xgapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 2201672 seqs, 1661799599 residues

Total number of hits satisfying chosen parameters: 4403344

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US0987755/runat_08122003_103436/app_query.fasta_1.391
-DB=Published Applications NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US0987755@cgn.1.1.221@runat_08122003_103436_18336
-NCPU=6 -ICPU=3 -NO MAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match Length	ID	Description
------------	-------------	--------------	----	-------------

1	1319	100.0	3202	11	US-09-987-755-1	Sequence 1, Appli
2	1319	100.0	3256	15	US-10-103-313-130	Sequence 33, Appl
3	1314	99.6	3172	15	US-10-103-313-33	Sequence 1, Appli
4	1291	97.9	3727	13	US-10-067-482-1	Sequence 133, App
5	1278	96.9	3707	15	US-10-103-313-133	Sequence 256, App
6	1263	95.8	810	15	US-10-103-313-256	Sequence 2, Appli
7	193.5	14.7	1898	9	US-09-768-826-22	Sequence 8, Appli
8	193.5	14.7	3805	13	US-10-247-671-8	Sequence 541, App
9	192	14.6	2920	11	US-09-866-050A-541	Sequence 541, App
10	192	14.6	2920	15	US-10-152-661-541	Sequence 68, Appli
11	186.5	14.1	2133	9	US-09-925-302-68	Sequence 1, Appli
12	185.5	14.1	723	15	US-10-207-791-1	Sequence 4, Appli
13	185.5	14.1	2376	9	US-09-938-885A-4	Sequence 3, Appli
14	185.5	14.1	2376	15	US-10-207-791-3	Sequence 2, Appli
15	179.5	13.6	869	9	US-09-938-885A-2	Sequence 145, App
16	177.5	13.5	1919	9	US-09-925-301-145	Sequence 24, Appli
17	166	12.6	1297	9	US-09-768-826-24	Sequence 307, App
18	166	12.6	2272	11	US-09-946-374-307	Sequence 307, App
19	166	12.6	2272	13	US-10-015-387A-307	Sequence 307, App
20	166	12.6	2272	13	US-10-006-130A-307	Sequence 345, App
21	166	12.6	2272	13	US-10-199-672-345	Sequence 307, App
22	166	12.6	2272	13	US-10-006-172A-307	Sequence 345, App
23	166	12.6	2272	13	US-10-187-749-345	Sequence 345, App
24	166	12.6	2272	13	US-10-194-457-345	Sequence 345, App
25	166	12.6	2272	13	US-10-184-642-345	Sequence 345, App
26	166	12.6	2272	13	US-10-196-747-345	Sequence 307, App
27	166	12.6	2272	13	US-10-015-392A-307	Sequence 307, App
28	166	12.6	2272	13	US-10-017-253A-307	Sequence 345, App
29	166	12.6	2272	13	US-10-173-689-345	Sequence 345, App
30	166	12.6	2272	13	US-10-173-690-345	Sequence 345, App
31	166	12.6	2272	13	US-10-173-691-345	Sequence 345, App
32	166	12.6	2272	13	US-10-173-692-345	Sequence 345, App
33	166	12.6	2272	13	US-10-173-694-345	Sequence 345, App
34	166	12.6	2272	13	US-10-173-698-345	Sequence 345, App
35	166	12.6	2272	13	US-10-173-699-345	Sequence 345, App
36	166	12.6	2272	13	US-10-173-707-345	Sequence 345, App
37	166	12.6	2272	13	US-10-173-707-345	Sequence 345, App
38	166	12.6	2272	13	US-10-174-583-345	Sequence 345, App
39	166	12.6	2272	13	US-10-174-587-345	Sequence 345, App
40	166	12.6	2272	13	US-10-174-589-345	Sequence 345, App
41	166	12.6	2272	13	US-10-174-591-345	Sequence 345, App
42	166	12.6	2272	13	US-10-175-736-345	Sequence 345, App
43	166	12.6	2272	13	US-10-175-742-345	Sequence 345, App
44	166	12.6	2272	13	US-10-175-744-345	Sequence 345, App
45	166	12.6	2272	13	US-10-175-745-345	Sequence 345, App

ALIGNMENTS

RESULT 1

US-09-987-755-1
; Sequence 1, Application US/0987755
; Publication No. US2003002312A1
; GENERAL INFORMATION:
; APPLICANT: Kunsch et al
; TITLE OF INVENTION: Human Hepatoma-Derived Growth Factor-2
; FILE REFERENCE: PF198D1C1
; CURRENT APPLICATION NUMBER: US/09/987,755
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: 09/263,625
; PRIOR FILING DATE: 1999-03-05
; PRIOR APPLICATION NUMBER: 08/464,600
; PRIOR FILING DATE: 1995-06-05
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 3202
; TYPE: DNA
; ORGANISM: human
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (34)..(780)
; OTHER INFORMATION:

US-09-987-755-1

```

Alignment Scores:
Pred. No.:      1,656-134      Length:      3202
Score:          1319.00      Matches:      249
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    100.00%      Indels:      0
DB:             11      Gaps:        0

```

US-09-987-755-2 (1-249) X US-09-987-755-1 (1-3202)

Qy	1	MetAlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTrpPro	20
Db	34	ATGGCGGCTGTGAGTCTCGGCTCGGCACTTGTGTGGGGAACTCGCCGATATCTT	93
Qy	21	ProTrpProGlyLysIleValAsnProProLysAspLeuLysValProArgGlyLysLys	40
Db	94	CCTTGGCCAGGAAGATGTATATCCCAAGGACTTGAAGAACTTCGGGAAGAA	153
Qy	41	CysPhePheValLysPhePheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeu	60
Db	154	TGCTCTTTGTGCAATTTTGGRAACAGCAATCATGCTTGGATCAAAGTGAACAGCTG	213
Qy	61	LysProTyrHisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGln	80
Db	214	AAGCCATATCATGCTCATAAAGAGAAATGATAAAATTTAACAGGGTAAACGATTTCCAG	273
Qy	81	GlnAlaValAspAlaValGluPheLeuArgAlaLysGlyLysAspGlnThrSer	100
Db	274	CAAGCGGTAGATGCTGTGAAAGATTCTCAGGAGAGCCAAAGGAAGACCAAGCTCA	333
Qy	101	SerHisAsnSerSerAspAspLysAsnArgAsnSerSerGluGluArgSerArgPro	120
Db	334	TCCACAAATTTCTTGATGACAAAGATCGACGTAAATCCAGTGAGGAGAGAGTAGGCCA	393
Qy	121	AsnSerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMet	140
Db	394	AACTCAGGTGATGAGAAGCGCAAACTTAGCCCTGTCTGAAGGAGAGGTGAAGAAGACATG	453
Qy	141	GlyGluGlyLysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeu	160
Db	454	GGAGAAGGAAGAAGAGGGTGTCTTCAGGCTCTTCAGAGAGAGGCTCCAATCCCTCTGTG	513
Qy	161	LysArgAlaGlnGluGlnSerProArgLysArgGlyArgProProLysAspGluLysAsp	180
Db	514	AAAAGAGCCCAAGACAAAGTCCCGGAAGCGGGTGGGCCCCCAAGAGATGAGAAGAT	573
Qy	181	LeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPhe	200
Db	574	CTCACCATCCCGAGTCTAGTACCGTGAAGGGGATGATGGCCGACCGCATGGCGCGGTTT	633
Qy	201	LysTrpGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisPhe	220
Db	634	AAATGGACGCCAACCGAAGCGAGCGCTGTTAAAGATGCATCTCTCATTTCCCATCATTT	693
Qy	221	LeuLeuSerGlnThrGluLysProAlaValCysTrpGlnAlaIleThrLysLysLeuLys	240
Db	694	CTGTCTAAGCCAAACAGAGAGCCAGCTGTCTGTTCACGACCATCAGAGAAGATTGAA	753
Qy	241	IleCysGluAspLeuLeuProArg	249
Db	754	ATATGTGAAGACCTCTCTTCTCTTAGG	780

RESULT 2

```

US-10-103-313-190
;
; Sequence 190, Application US/10103313
; Publication No. US20030082759A1
;
; GENERAL INFORMATION:
;
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJ207C1
; CURRENT APPLICATION NUMBER: US/10/103.313
;

```

```

; CURRENT FILING DATE: 2002-03-12
; NUMBER OF SEQ ID NOS: 653
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 190
; LENGTH: 3256
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3220)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-033-190

```

Alignment Scores:	
Pred. No.:	1.69e-134
Score:	1319.00
Length:	3256
Matches:	249
Conservative:	0
Mismatches:	0
Best Local Similarity:	100.00%
Query Match:	100.00%
Indels:	0
Gaps:	0
DB:	15

US-09-987-755-2 (1-249) x US-10-103-313-190 (1-3256)

QY	1	MetAlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrPro	20
DB	34	ATGGCGGCTGTGAGTCTCGGCTCGGCGACTTGTGTGGGGGAAACTCGCGCGATATCCT	93
QY	21	ProTrpProGlyLysLeuValAsnProProLysAspLeuLysLysProArgGlyLysLys	40
DB	94	CCTTCGGCAGGAAGATTGTTAATCCAAAGAGACTTGAAGAAACCTTCGCGGAAGAAA	153
QY	41	CysPhePheValLysPhePheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeu	60
DB	154	TGCTCTCTTGTGAAATTTTGGAAACAGAGATCATCGCTGGATCAAAGTGGAAACAGCTG	213
QY	61	LysProTyrHisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGln	80
DB	214	AAAGCATATCATGCTCATAAAGAGAAATGATAAAATTAACAAGGGTAAACCATTTCCAG	273
QY	81	GlnAlaValAspAlaValGluPheLeuArgArgAlaLysGlyLysAspGlnThrSer	100
DB	274	CAAGCGGTAGATGCTGTGCAAGAGTTCTCAGGAGGCCAAAGGGGAAAGCCAGACGTCA	333
QY	101	SerHisAsnSerSerAspAspLysAsnArgArgAsnSerSerGluGluArgSerArgPro	120
DB	334	TCCCACAATTTCTTGATGACAGAANTCGAGCTAATTCAGTCAGGAGAGAGTAGGCCA	393
QY	121	AsnSerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMet	140
DB	394	AACTCAGGTGATGAGAGCGCAACTTAGCTGTCTGAAGGGAGGTTGAAGAAGAACATG	453
QY	141	GlyGluGlyLysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeu	160
DB	454	GGAGAAGGAAAGAGAGGGTGTCTTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTG	513
QY	161	LysArgAlaGlnGluGlnSerProArgLysArgGlyArgProProLysAspGluLysAsp	180
DB	514	AAAAGAGCCCAAGAGCAAAAGTCCCGAAGCGGGTCCGGCCCCCAAGAGATGAGAAGGAT	573
QY	181	LeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPhe	200
DB	574	CTCACCATCCCGGAGTCTAGTACCGTGAAGGGGATGATGCCGAGCCGATGGCCGGTTT	633
QY	201	LysTrpGlnProThrAlaSerGluProValLysAspAlaAspProHisHisLysPhe	220
DB	634	AAATGGCAGCAACCGCAAGCGAGCTGTAAAGATGCAGATCCTCATTTTCCATCATTTTC	693
QY	221	LeuLeuSerGlnThrGluLysProAlaValCysTyrGlnAlaIleThrLysLysLeuLys	240
DB	694	CTGCTAAAGCCAAACAGAGAGCCAGCTGTCTGTGTACCAGGCAATCAGGAAGAAGTTGAA	753
QY	241	IleCysGluAspLeuLeuLeuProArg	249

```
Db 754 ATATGTGAAGACCTCTTCTTCTTAGG 780
|||||
RESULT 3
US-10-103-313-33
; Sequence 33, Application US/10103313
; Publication No. US20030082758A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P207C1
; CURRENT APPLICATION NUMBER: US/10/103,313
; CURRENT FILING DATE: 2002-03-12
; NUMBER OF SEQ ID NOS: 653
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 33
; LENGTH: 3172
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-103-313-33
Alignment Scores:
Pred. No.: 5,77e-134 Length: 3172
Score: 1314.00 Matches: 248
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.62% Indels: 0
DB: 15 Gaps: 0
US-09-987-755-2 (1-249) x US-10-103-313-33 (1-3172)
QY 2 AlaAlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrProPro 21
Db 9 GCGGCTGTGAGTCTCGGCTCGCGACTTGTGTGGGGAACTCGGCCGATATCTCT 68
QY 22 TrpProGlyLysLeuValAsnProProLysAspLeuLysProArgGlyLysCys 41
Db 69 TGGCCAGAAAGATTGTTAATCCACCAAGGACTTGAAGAAACCTCGCGGAAAGAAATGC 128
QY 42 PhePheValLysPheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeuLys 61
Db 129 TTCTTGTGAAATTTTGGAAACAGAGATCATGCTGGATCAAAAGTGGAACTGAAG 188
QY 62 ProTyrHisAlaHisLysGluGluMetIleLysLysLysLysGlyLysArgPheGlnGln 81
Db 189 CCATATCATGCTCATAAAGAGGAATGATATAAATTAACAGGGTAAACGATTCACGAA 248
QY 82 AlaValAspAlaValGluGluPheLeuArgArgAlaLysGlyLysAspGlnThrSerSer 101
Db 249 GCGGTAGATGCTGTGGAAGAGTTCTCTCAGGAGAGCCAAAGGAAAGACGATCATCC 308
QY 102 HisAsnSerSerAspLysAsnArgArgAsnSerSerGluGluArgSerArgProAsn 121
Db 309 CACAATTTCTTGTGATCAAGAAATCGACGTAATTCAGTGAGGAGAGATAGGCCAAAC 368
QY 122 SerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValLysAsnMetGly 141
Db 369 TCAGGTATGAGAGCCCAAACTTAGCTGCTGTAAGGAGAGGTGAAGAGAAATCGGA 428
QY 142 GluGlyLysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeuLys 161
Db 429 GAAGAAAGAGAGGGGTCTTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTGAAA 488
QY 162 ArgAlaGlnGlnSerProArgLysArgGlyArgProProLysAspGluLysAspLeu 181
Db 489 AGAGCCCAAGAGCAAAAGTCCCGAAGCGGGTTCGGCCCAAGAGGATGAGAGGATCTC 548
QY 182 ThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPheLys 201
Db 549 ACCATCCGGAGTCTAGTACCGTAGGGGATGATGCCGAGCCGATGGCCGCTTTAAA 608
QY 202 TrpGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisPheLeu 221
|||||
Db 609 TGGCAGCCCAACCGCAGCGAGCTGTAAAGATGCGAGATCCTCATTTCCATCTCTG 668
QY 222 LeuSerGlnThrGluLysProAlaValCysTyrGlnAlaIleThrLysLysLeuLysIle 241
Db 669 CTAAGCCCAACAGAGAGCGAGCTGTCTGTACAGGCAATCAGAGAGATTGAAATA 728
QY 242 CysGluAspLeuLeuProArg 249
Db 729 TGTGAAGACCTCTCTTCTCTAGG 752
RESULT 4
US-10-067-482-1
; Sequence 1, Application US/10067482
; Publication No. US20030148407A1
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies, Inc.
; TITLE OF INVENTION: Angiogenesis Dehydrogenase Gene
; FILE REFERENCE: IU 102 K1
; CURRENT APPLICATION NUMBER: US/10/067,482
; CURRENT FILING DATE: 2002-02-07
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 1
; LENGTH: 3727
; TYPE: DNA
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (28)..(1689)
; OTHER INFORMATION:
US-10-067-482-1
Alignment Scores:
Pred. No.: 2,37e-131 Length: 3727
Score: 1291.00 Matches: 243
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.59% Mismatches: 0
Query Match: 97.88% Indels: 0
DB: 13 Gaps: 0
US-09-987-755-2 (1-249) x US-10-067-482-1 (1-3727)
QY 1 MetAlaAlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrPro 20
Db 28 ATGCGGCTGTGAGTCTCGGCTCGCGACTTGTGTGGGGAACTCGGCCGATATCTCT 87
QY 21 ProTrpProGlyLysLysLeuValAsnProProLysAspLeuLysProArgGlyLysLys 40
Db 88 CCTTGGCCAGAAAGATTGTTAATCCACCAAGGACTTGAAGAAACCTCGCGGAAAGAAA 147
QY 41 CysPhePheValLysPheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeu 60
Db 148 TGCTTCTTGTGAAATTTTGGAAACAGAGATCATGCTGGATCAAAAGTGGAACTG 207
QY 61 LysProTyrHisAlaHisLysGluGluMetIleLysLysLysLysGlyLysArgPheGln 80
Db 208 AACCCATATCATGCTCATAAAGAGGAATGATATAAATTAACAAAGGGTAAACGATTCAG 267
QY 81 GlnAlaValAspAlaValGluGluPheLeuArgArgAlaLysGlyLysAspGlnThrSer 100
Db 268 CAAGCGGTAGATGCTGTGGAAGAGTTCTCTCAGGAGAGCCAAAGGAAAGACGACGTCA 327
QY 101 SerHisAsnSerSerAspLysAsnArgArgAsnSerSerGluGluArgSerArgPro 120
Db 328 TCCCAAAATTTCTTCTGATGCAAGATTCGACGTAATTTCCAGTGAGGAGAGATGAGCCCA 387
QY 121 AsnSerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMet 140
Db 388 AACTCAGGTGATGAGAGCCCAAACTTAGCTGCTGTAAGGAGGATGAGAGAGACATG 447
QY 141 GlyGluGlyLysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeu 160
|||||
```

Db 448 GGAGAGAAAGAGAGAGGTGTCTTACGGCTCTTTCAGAGAGAGGCTCCAAATCCCTCTG 507
Qy 161 LysArgAlaGlnGluSerProArgLysArgGlyArgProProLysAspGluLysAsp 180
Db 508 AAAAGAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTCCGCCCCCAAGAGGATGAGAAGGAT 567
Qy 181 LeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPhe 200
Db 568 CTCACCATCCCGGAGTCTAGTACCGTGAAGGGGATGATGGCGGACCGATGGCCGGTCTT 627
Qy 201 LysTrpGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisPhe 220
Db 628 AAATGGCAGCAACCGCAAGCAGGCTGTAAAGATGAGATCTCTCATTTCCATCATTTTC 687
Qy 221 LeuLeuSerGlnThrGluLysProAlaValCysTrpGlnAlaIleThrLysLysLys 240
Db 688 CTGTCTAAGCCAAACAGAGAGCCAGCTGTCTGTACCAAGGCAATCACGAAGAAGTTGAA 747
Qy 241 IleCysGluAsp 244
Db 748 ATATGTGAAGAG 759
RESULT 5
US-10-103-313-133
; Sequence 133, Application US/10103313
; Publication No. US20030082758A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJ207C1
; CURRENT APPLICATION NUMBER: US/10/103,313
; CURRENT FILING DATE: 2002-03-12
; NUMBER OF SEQ ID NOS: 653
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 133
; LENGTH: 3707
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-103-313-133
Alignment Scores:
Pred. No.: 6,288-130 Length: 3707
Score: 1278.00 Matches: 240
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.59% Mismatches: 0
Query Match: 96.89% Indels: 0
DB: 15 Gaps: 0
US-09-987-755-2 (1-249) x US-10-103-313-133 (1-3707)
Qy 4 ValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyProProTrpPro 23
Db 20 GTGAGTCTCGCGCTCGCGACTTGTGTGGGGAACTCGGCGGATATCTCTTGGCCA 79
Qy 24 GlyIleValAsnProProLysAspLysLysProArgGlyLysCysPhe 43
Db 80 GGAAAGATTGTAAATCCCAAGAGACTTGAAGAAACCTCGCGAAGAAATGCTTCTTT 139
Qy 44 ValLysPhePheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeuLysProTy 63
Db 140 GTGAAATTTTGGNACAGAGATCATCGCTGGATCAAGTGAACACACTGAGCCATAT 199
Qy 64 HisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGlnAlaVal 83
Db 200 CATGCTCATAAGAGGAAATGATAAAATTAACAGGGTAAACGATTCAGCAAGCGGTA 259
Qy 84 AspAlaValGluGluPheLeuArgAlaLysGlyLysAspGlnThrSerSerHisAsn 103
Db 260 GATGCTGTGGAAGAGTTCTCTCAGGAGAGCCAAAGGGAAGACGACGTATCCCAAT 319
Qy 104 SerSerAspLysAsnArgAsnSerSerGluGluArgSerArgProAsnSerGly 123

Db 320 TCTTCTGATGACAAAGATTCGACGTAAATTCAGTCAGGAGAGAGTAGGCCAAACTCAGGT 379
Qy 124 AspGluLysArgLysLeuSerLeuSerGluGlyLysValLysAsnMetGlyGluGly 143
Db 380 GATGAGAAGCGCAAACTTAGCTGTCTGAAAGGGAAGGTGAAGAAACATGGGAGAAGGA 439
Qy 144 LysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeuLysArgAla 163
Db 440 AAGAAAGAGGTGTCTTTCAGGCTCTTTCAGAGAGAGGCTCCAAATCCCTCTGAAAGAGCC 499
Qy 164 GlnGluGlnSerProArgLysArgGlyArgProProLysAspGluLysAspLeuThrIle 183
Db 500 CAAGAGCAAAAGTCCCGGAGCGGGTCCGCCCCCAAGGATGAGAAGGATCTCACCATC 559
Qy 184 ProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPheLysTrpGln 203
Db 560 CCGAGGTCTTAGTACCGTGAAGGGGATGATGCCCGGACCGATGGCCGGTCTTAAATGGCAG 619
Qy 204 ProThrAlaSerGluProValLysAspAlaAspProHisPheHisPheLeuLeuSer 223
Db 620 CCAACCGCAAGCGGCTCTTAAAGATGAGATCTCTCATTTCCATCTTCTGCTAAGC 679
Qy 224 GlnThrGluLysProAlaValCysTrpGlnAlaIleThrLysLysLeuLysIleCysGlu 243
Db 680 CAAACAGAGAAGCCAGCTGTCTGTACAGGCAATCACGAAGAGTTGAAATATGTGAA 739
Qy 244 Asp 244
Db 740 GAG 742
RESULT 6
US-10-103-313-256
; Sequence 256, Application US/10103313
; Publication No. US20030082758A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJ207C1
; CURRENT APPLICATION NUMBER: US/10/103,313
; CURRENT FILING DATE: 2002-03-12
; NUMBER OF SEQ ID NOS: 653
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 256
; LENGTH: 810
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (12)_feature
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (43)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-103-313-256
Alignment Scores:
Pred. No.: 3,9e-129 Length: 810
Score: 1263.00 Matches: 238
Percent Similarity: 99.17% Conservative: 1
Best Local Similarity: 98.76% Mismatches: 2
Query Match: 95.75% Indels: 0
DB: 15 Gaps: 0
US-09-987-755-2 (1-249) x US-10-103-313-256 (1-810)
Qy 4 ValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyProProTrpPro 23
Db 2 GTGAGTCTCGCGCTCGCGACTTGTGTGGGGAACTCGGCGGATATCTCTTGGCCA 61
Qy 24 GlyIleValAsnProProLysAspLysLysProArgGlyLysCysPhe 43
Db 62 GGAAAGATTGTAAATCCCAAGAGACTTGAAGAAACCTCGCGAAGAAATGCTTCTTT 121


```

Qy 44 ValLysPheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeuLysProTyr 63
Db 122 GTGAAATTTTGGACACAGAGATCATGCTGGATCAAGTGGAAACAGCTGAAGCCATAT 181
Qy 64 HisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGlnAlaVal 83
Db 182 CATGCTCATAAAGAGGAAATGATAAAATTAACAAAGGTAAACGATTCCAGCAAGCGGTA 241
Qy 84 AspAlaValGluGluPheLeuArgAlaLysGlyLysAspGlnThrSerSerHisAsn 103
Db 242 GATGCTGTGAGAGTTCTTCAGAGAGCCAAAGGGAAGNCCAGAGCTATCCCAAT 301
Qy 104 SerSerAspLysAsnArgArgAsnSerSerGluGluArgSerArgProAsnSerGly 123
Db 302 TCTTCTGATCAAGAATCGACGTAATTCAGTGAGGAGAGAGTAGGCCAACTCAGGT 361
Qy 124 AspGluLysArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMetGlyGluGly 143
Db 362 GATGAGAGGCGCAACTTAGCTCTCTGAAGGGAAGGTGAAGAAACATGGGGAAGGA 421
Qy 144 LysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeuLysArgAla 163
Db 422 AAGAAGAGGGTGCTTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTGAAAGAGGCC 481
Qy 164 GlnGluGlnSerProArgLysArgGlyArgProProLysAspGluLysAspLeuThrIle 183
Db 482 CAAGAGCAAGTGTCCCGAAGCGGGTGGGCCCCCAAGAGATGAGAGGATCTCAACATC 541
Qy 184 ProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPheLysTrpGln 203
Db 542 CCGAGGTCTAGTACCGTGAAGGGGATGATGGCCGAGCCGATGGCCGGCTTTAAATGGCAG 601
Qy 204 ProThrAlaSerGluProValLysAspAlaAspProHisPheHisHisPheLeuLeuSer 223
Db 602 CCAACCGAAGCGAGCGCTGTAAAGATGCAGATCTCATTTCCATCATTTCTCTCTAAGC 661
Qy 224 GlnThrGluLysProAlaValCysTrpGlnAlaIleThrLysLysLeuLysIleCysGlu 243
Db 662 CAACACAGAGAGCCAGCTGTCTGTACAGGCAATCAGAGAGATTTGAAATATGTGAA 721
Qy 244 Asp 244
Db 722 GAG 724

RESULT 7
US-09-768-826-22
; Sequence 22, Application US/09768826
; Patent No. US2002012966A1
; GENERAL INFORMATION:
; APPLICANT: Shi et al.
; TITLE OF INVENTION: 18 human secreted proteins
; FILE REFERENCE: PF512P1
; CURRENT APPLICATION NUMBER: US/09/768,826
; CURRENT FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: PCT/US00/22350
; PRIOR FILING DATE: 2000-08-15
; PRIOR APPLICATION NUMBER: 60/148,759
; PRIOR FILING DATE: 1999-08-16
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 1898
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-768-826-22

Alignment Scores:
Pred No. : 2,45e-11 Length: 1898
Score: 193.50 Matches: 67
Percent Similarity: 45.26% Conservative: 38
Best Local Similarity: 28.88% Mismatches: 101
Query Match: 14.67% Indels: 27

```

```

Db: 9 Gaps: 7

```

```

US-09-987-755-2 (1-249) x US-09-768-826-22 (1-1898)
Qy 7 ArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrProTrpProGlyLysIle 26
Db 177 AAAGCGGCGACCTGGTCTTCGCCAAGATGAAGGGCTACCGCACTGGCGGCGCGGATT 236
Qy 27 ValAsnProProLysLysLeuLysLysLysLysLysLysLysLysLysLysLysLys 46
Db 237 GATGAACCTCCAGAGGGCGCTGTGAAGCCTCCAGCAACCAAG--TATCTATCTCTTT 293
Qy 47 PheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeuLysProTyrHisAlaHis 66
Db 294 TTTGGCACCCATGCAATTTCTAGTGTCCCAAGACCTTTTCCATATAAGGAGTAC 353
Qy 67 LysGluGluMetIleLysIleAsnLysGlyLysArgPheGlnGlnAlaValAspAlaVal 86
Db 354 AAAGACAAAGTTTGGAAAGTCAAAACCAACGGAAGAGATTTAACGAAGGATTGTGGAAATA 413
Qy 87 Glu-----GluPheLeuArgAlaLysGlyLysAspGlnThrSerSer 101
Db 414 GAAATAATACCCAGAGGTAAAGTTTACTGGCTACCGCAATTTCAGCAACAGAGCTCTTCA 473
Qy 102 HisAsnSerSerAspLysAsnArgArgAsnSerSerGluGluArgSerArgProAsn 121
Db 474 GAAACTGAGGGAGAGGTGGAATACTGTCAGATGCAAGCAGTGAG-----GAA 521
Qy 122 SerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValLys--LysAsnMet 140
Db 522 GAAGGTGATAGATGAGAGAA-----GATGAAAGGCAAGAAAGAAATGAA 569
Qy 141 GlyGluGlyLysLysArgValSerSerGlySerSerGluArgGlySerLys-SerProLe 160
Db 570 AAAGCAGGCTCAAAACGGAAGAAAGTCAATATCTTCAAGAAATCTCTTAACAGTCCC-- 627
Qy 160 uLysArgAlaGlnGlnSerProArgLysArgGlyArgProProLysAspGluLysAs 180
Db 628 -CGAAATCTCCAGGAGATGAAGATGACAAAGACTGCAAGAGAGAGAAACAAAGAGCAGC 686
Qy 180 pLeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPh 200
Db 687 TCTGAGGGTGGAGATCGGGCAACGACACAAAGAAACCAACTT----- 729
Qy 200 eLysTrpGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisPh 220
Db 730 -----CACACTTGCAAGAAACCAAGTGAAGGGACCTTAACCTACCATATGATGATG 779
Qy 220 eLeuLeuSerGlnThrGluLysProAlaValCys 231
Db 780 CATATTAAAGAGAAACCCAC-AAGAAGGTTATATGT 812

```

```

RESULT 8
US-10-247-671-8
; Sequence 8, Application US/10247671
; Publication No. US20030194721A1
; GENERAL INFORMATION:
; APPLICANT: Mikita, Thomas
; APPLICANT: Shiffman, Dov
; APPLICANT: Porter, Gordon, J.
; APPLICANT: Kaser, Matthew R.
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED FOAM CELLS
; FILE REFERENCE: PA-0050 US
; CURRENT APPLICATION NUMBER: US/10/247,671
; CURRENT FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: 60/323,784
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 186
; SOFTWARE: PERL Program
; SEQ ID NO 8
; LENGTH: 3805
; TYPE: DNA
; ORGANISM: Homo sapiens

```

```

: FEATURE:
: NAME/KEY: misc feature
: OTHER INFORMATION: Incyte ID No. US20030194721A1 2676869CB1
: FEATURE:
: NAME/KEY: unsure
: LOCATION: 3788
: OTHER INFORMATION: a, t, c, g, or other
: US-10-247-671-8

```

Alignment Scores:	
Pred. No.:	66-11
Score:	193.50
Percent Similarity:	45.28%
Best Local Similarity:	28.88%
Query Match:	14.67%
DB:	13
Length:	3805
Matches:	67
Conservative:	38
Mismatches:	101
Indels:	27
Gaps:	7

US-09-987-755-2 (1-249) x US-10-247-671-8 (1-3805)

Qy	7	ArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrProProTrpProGlyLysIle	26
Db	151	AAAGCGGGCAGCTGGTCTTCGCCAAGATGAAGGGCTACCGGCATCTGGCGGGCCGGATT	210
Qy	27	ValAsnProProLysAspLeuLysAspProArgGlyLysLysCysPhePheValLysPhe	46
Db	211	GATGAACCTCCAGAGGGCGCTGGAAGCTTCAGCAAAACAG---TATCTATCTTCTTT	267
Qy	47	PheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeuLysProTyrHisAlaHis	66
Db	268	TTTGCCACCCTCAAACTGCATTTCTAGGTCCTCCAAAGACCTTTTTCATATAAGGAGTAC	327
Qy	67	LysGluGluMetIleLysIleAsnLysGlyLysArgPheGlnGlnAlaValAspAlaVal	86
Db	328	AAAGACAACTTTGGAAAGTCAACCAACCGAAAGCAATTTAACGAAAGGATTTGTGGAAATA	387
Qy	87	Glu-----GluPheLeuArgAlaLysGlyLysAspGlnThrSerSer	101
Db	388	GAATAATAACCCAGGAGTAAAGTTTACTGGCTACCAGGCAATTCAGCAACAGAGCTCTTCA	447
Qy	102	HisAsnSerSerAspAspLysAsnArgAsnSerSerGluGluArgSerArgProAsn	121
Db	448	GAACCTGAGGAGAGAGGTGGAAATACTGCAGATCAACAGCAGTGAG-----GAA	495
Qy	122	SerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValLys---LysAsnMet	140
Db	496	GAAGGTGATAGTAGAAGAA-----GATGNAAGAGCCAAAGAAAGATGAA	543
Qy	141	GlyGluGlyLysLysArgValSerSerGlySerSerGluArgGlySerLys-SerProLeu	160
Db	544	AAAGCAGGCTCAAAACGGAAAAAGTCATATACTTCAAGAAATCCTCTAAACAGTCCC--	601
Qy	160	uLysArgAlaGlnGluGlnSerProArgLysArgGlyArgProProLysAspGluLysAs	180
Db	602	-CGAAATCTCCAGGAGATGAAGATGACAAAGACTGCAAGAGGAGGAAACAAAGCAGC	660
Qy	180	pLeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPh	200
Db	661	TCTAGGGTGGAGATCGCGGCAACGACACAAAGAAACACAACTT-----703	
Qy	200	eLysTrpGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisPh	220
Db	704	-----CAGACTTCAGAAACACAGTGAAGGACCTTAACCTACCTAATGAATGCTG	753
Qy	220	eLeuLeuSerGlnThrGluLysProAlaValCys	231
Db	754	CATATTAAGAGAAACCAAC-AAGAAGTTTATATGT	786

RESULT 9

US-09-866-050A-541
; Sequence 541, Application US/09866050A
; Publication No. US2003004071A1
; GENERAL INFORMATION:
; APPLICANT: watson, James D.

```
Db 908 CTGCATATTGAGAGAAACCAAGAGTTAACTGTTGATTGCTGATTCTTGGATTG 967
Qy 228 -----ProAlaValCysTyrGlnAlaIleThrLysLeuLysIleCys 242
Db 968 ATATGAACCAACAGTCTTT-----GTTGTCACTGACAAAGAGCCCGAGTGTGT 1013

RESULT 10
US-10-152-661-541
; Sequence 541, Application US/10152661
; Publication No. US2003002835A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishanand D.
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods for Their Use
; FILE REFERENCE: 11000.1011c5
; CURRENT APPLICATION NUMBER: US/10/152.661
; CURRENT FILING DATE: 2002-05-20
; PRIOR APPLICATION NUMBER: 09/866,050
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/221,232
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: 60/206,650
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: 09/312,283
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: PCT/NZ99/00051
; PRIOR FILING DATE: 1999-04-29
; PRIOR APPLICATION NUMBER: 09/188,930
; PRIOR FILING DATE: 1998-11-09
; PRIOR APPLICATION NUMBER: 09/069,726
; PRIOR FILING DATE: 1998-04-29
; NUMBER OF SEQ ID NOS: 725
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 541:
; LENGTH: 2920.
; TYPE: DNA
; ORGANISM: Rat
US-10-152-661-541

Alignment Scores:
Pred. No.: 6.23e-11 Length: 2920
Score: 192.00 Matches: 69
Percent Similarity: 43.41% Conservative: 43
Best Local Similarity: 26.74% Mismatches: 101
Query Match: 14.56% Indels: 45
DB: 15 Gaps: 9

US-09-987-755-2 (1-249) x US-10-152-661-541 (1-2920)

Qy 7 ArgLeuGlyAspLeuValTyrGlyLysLeuGlyArgTyrProTyrProGlyLysIle 26
Db 311 AAACGGGAGACCTGGTTCGCCAAATGAAGGGCTACCGCACTGGCCGCCGGATT 370
Qy 27 ValAsnProProLysAspLeuLysLysProArgGlyLysCysPheValLysPhe 46
Db 371 GATGAACCTCCAGAGGGCGCGTGAAGCCTCCAGCAACAAAG---TACCCTATCTCTTT 427
Qy 47 PheGlyThrGluAspHisAlaTyrIleLysValGluGlnLeuLysProTyrHisAlaHis 66
Db 428 TTCGGAACCCATGAACCTGCAATTTCTAGGTCTTAAGACCTTTTCCCATATAAGGAATAC 487
Qy 67 LysGluGluMetIleLysIleAsnLysGlyLysArgPheGlnGlnAlaValAspAlaVal 86
Db 488 AAAGACAAGTTTGGAAAGTCAAAACCGGAAGAGATTTAATGAAGGATTATGGGAATTT 547
Qy 87 Glu-----GluPheLeuArgAlaLysGlyLysAspGlnThrSerSer 101
Db 87 Glu-----GluPheLeuArgAlaLysGlyLysAspGlnThrSerSer 101
```

```
Db 548 GAAAAATAATCCAGGAGTGAATTTACTGGGTACCAGACAAATTCCAGCAACAGAGCTTTCA 607
Qy 102 HisAsnSerSerAspAspLysAsnArgAsnSerSerGluGluArgSerArgProAsn 121
Db 608 GAAACTGAGGAGAGGAGGACACTGAGATGCAAGCAGTGAAG-----GAA 655
Qy 122 SerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMetGly 141
Db 656 GAAGGTGAC-----AGAGTAGAAGATGGA 679
Qy 142 GluGlyLysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeuLys 161
Db 680 AAAGGCAAGAGAAAGAAATGAAAGAGGCTCAAAACGGGAAAAAGCTCTACCTTCAAG 739
Qy 162 ArgAlaGlnGluGlnSerProArgLysArgGlyArgProProLysAspGlyLysAspLeu 181
Db 740 AAGTCTTCTAAACAGTCCCGGAAATCTCCAGGA-----GACGAAGATGACAAAGACTGC 793
Qy 182 ThrIleProGlu-----SerSerThrValLysGly-MetMetAlaGlyProMetAlaAl 199
Db 794 AAAGAAGAGGAGAGAAACAAAGCAGCTCTGAGGGCGGAGATGCTGGCA-----ATGACACA 847
Qy 199 apheLysTyrGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHis 219
Db 848 AGAAACACGACTTCAGACTTGCAGAAAGCGGTGAAGGGGACCTAACTACCGTAATGAATG 907
Qy 219 sPheLeuLeuSerGlnThrGluLys----- 227
Db 908 CTGCATATTGAGAGAAACCAACAGAGTTAACTGTTGATTGCTGATTCTTGGATTG 967
Qy 228 -----ProAlaValCysTyrGlnAlaIleThrLysLysLeuLysIleCys 242
Db 968 ATATGAACCAACAGTCTTT-----GTTGTCACTGACAAAGAGCCCGAGTGTGT 1013

RESULT 11
US-09-925-302-68
; Sequence 68, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: FA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 68
; LENGTH: 2133
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-302-68

Alignment Scores:
Pred. No.: 1.67e-10 Length: 2133
Score: 186.50 Matches: 64
Percent Similarity: 43.72% Conservative: 30
Best Local Similarity: 29.77% Mismatches: 100
Query Match: 14.14% Indels: 21
DB: 9 Gaps: 8

US-09-987-755-2 (1-249) x US-09-925-302-68 (1-2133)

Qy 7 ArgLeuGlyAspLeuValTyrGlyLysLeuGlyArgTyrProTyrProGlyLysIle 26
Db 49 AAATGGGGAGACCTGGTGTTCGCCAAGATGAAGGGCTACCCACACTGGCCGCCGGATT 108
Qy 27 ValAsnProProLysAspLeuLysLysProArgGlyLysCysPheValLysPhe 46
Db 109 GACGAGATGCTGAGGCTGCCGTGAATCAACAGCCCAACAA---TACCAAGTCTTTT 165
```


TELEPHONE: 415-855-0555

TEL/FAX: 415-845-4166

TELEX: <linknow>

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS.

ANCE CHARACTERISTICS:
LENGTH: 2276 base pairs

LENGTH: 2376 BASE PAIRS
TYPE: uncloned cDNA

TYPE: NUCLEIC ACID
TRANSCRIPTION: SIGNAL

SIRANDEDNESS: single
TOROIDAL: 1:1:1

INVERTED COMET TOPOLOGY: I

IMMEDIATE SOURCE:

LIBRARY: GenB

CLONE: 598956

SEQUENCE:

Alignment Scores:	
Pred. No.:	2.47e-10
Score:	185.50
Percent Similarity:	43.72%
Best Local Similarity:	29.77%
Query Match:	14.06%
DB:	9
Length:	2376
Matches:	64
Conservative:	30
Mismatches:	100
Indels:	21
Gaps:	8

US-09-987-755-2 (1-249) X US-09-938-885A-4 (1-2376)

QY	7	ArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrProProTrpProGlyLysIle	26
Db	346	AAATCGGGGACCTTGTTCCGCAAGATGAAGGGCTACCCACACATGGCGCGCCGGATT	405
QY	27	ValAsnProProLysAspLeuValLysValProArgGlyLysLysCysPhePheValLysPhe	46
Db	406	GACGAGATGCCTGAGGCTGCCGTGAATCAACACCCAAACAA---TACCAGTCTTTTTT	462
QY	47	PheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeuLysProTyrHisAlaHis	66
Db	463	TTTCGGGACCCACGACGCGCATCTCTGGGCCCAAGACCTCTTCCCTACGAGGAATCC	522
QY	67	LysGluGluMetIleLysIleAsnLysGlyLysArgPheGlnGlnAlaValaAspAlaVal	86
Db	523	AAAGGAGAGATTGGCAAGCCCAACAGAGGAAGGGTTACAGCGGGGCTGTGGGAGATC	582
QY	87	GluGluPheLeuArgAlaLysGlyLysAspGlnThrSerSerHisAsnSerSerAsp	106
Db	583	GAGAAC-----AACCTACTGTCAAGGCTTCGGGTATCATGCTCTCCACG	627
QY	107	AspLysAsnArgArgAsnSerSerGluGluArgSerArgProAsnSerGlyAspGluLys	126
Db	628	AAAAAGAGCTGTGTGGAAGAGCCTGAACACAGAGCCCGAAGCTGCAGAGGGTGACGGTGAT	687
QY	127	ArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMetGlyGluGlyLysLysArg	146
Db	688	AAGAG--GGGAATGCAGAGCGCAGCAGCACAG-----GAAGGGAAGCTGGTC	735
QY	147	ValSerSerGlySerSerGluArgGlySerLysSerProLeuLysArgAlaGln-----	164
Db	736	ATTGATGAGCCAGCAGAGAAGAACAGAAAGAGCGCTTGAAGAGCAGCAGCGGAC	795
QY	165	-----GluGlnSerProArg-----LysArgGlyArgProProLysAspGluLysAsp	180
Db	796	TTGCTGGAGGACTCTCTTAACGTCCCAAGGAGCGCAAAACCCCTGAAGAGAGAGAG--	852
QY	181	LeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPhe	200
Db	853	-----AAGGAGCGACCACTTGAGGTTGAGAGGCCCTTCCTATGGAGGTGGAA	903
QY	201	LysTrpGlnProThrAlaSerGluProValLysAspAlaAspPro	215
Db	904	AAG-----AATAGCACCCCTCTGAGCCCGGCTCTGGCGGGGGCT	945

DECIII.T 1A

ИТС-10-207-701-3

US-10-20/-791-3
: gac:ccc ? bac:ccc 10/10000000

sequence 3, APPLICATION US/10201702013042381

: GENERAL INFORMATION:

: APPL.TCANT: pharmaDesign. Inc: Masaki Mori

TITLE OF INVENTION: A prediction method of

1. TITLE OF INVENTION: A PROCESS FOR PRODUCING A POLYMER
2. TITLE REFERENCE: PDP-0016

FILE REFERENCE: 121 0019
CURRENT APPLICATION NUMBER: IIS/10/207 791

CURRENT AFFILIATION NUMBER: 05/10/2017, 151
CURRENT FILING DATE: 2003-07-31

; CURRENT FILING DATE: 2002-07-31
 : NUMBER OF SEC TO NOS. 8

; NUMBER OF SEQ ID NOS: 6
 COUNTRY: Denmark Nov 21

```

; SOFTWARE: PALCIII VER. 2.1
; CEO TO NO ?

```

3 NO ID TO SEQ ;

; LENGTH: 2376
MUSE: 0117

; ;
TYPE: DNA
CONTENT: " " ;

ORGANISM: Homo sapiens.

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (316)..(1038)

Alignment Scores:	
Pred. No.:	2.47e-10
Score:	185.50
Percent Similarity:	43.72%
Best Local Similarity:	29.77%
Query Match:	14.06%
DR:	15
Length:	2376
Matches:	64
Conservative:	30
Mismatches:	100
Indels:	21
Gaps:	8

110 00 007 755 3 (1 240) * 118 10 207 701 3 (1 2275)

QY	7	ArgLeuGlyAspLeuValTrpGlyLysLeuIglyArgTyrProProTrpProGlyLysIle	26
Db	346	AAATCGCGGGACCTGGTGTTCGCCCAAGATGAAGGGCTACCCACACATGGCCGGCCGGATT	405
QY	27	ValAsnProProLysAspLeuLysLysProArgGlyLysLysCysPhePheValLysPhe	46
Db	406	GACGAGATGCTGAGGGTCGCGTGAATATCAACGCCAACAAA---TACCAAGTCTTTTT	462
QY	47	PheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeuLysProTyrHisAlaHis	66
Db	463	TTCGGACCCACGACGAGCGCATCTCTGGGCCCAAGACCTCTCCCTTACGAGGAATCC	522
QY	67	LysGluGluMetIleLysIleAsnLysGlyLysArgPheGlnGlnAlaValAspAlaVal	86
Db	523	AAGGAGAAGTTTGGCAGAGCCCAACAGAGGAAGGGTTTCAGCGGGGCTGTGGGAGATC	582
QY	87	GluGluPheLeuArgAlaLysGlyLysAspGlnThrSerSerHisAsnSerSerAsp	106
Db	593	GAGAAC-----AACCTTACTGTCAAGCTTCGGCTATCATGCTCTCCAG	627
QY	107	AspLysAsnArgArgAsnSerSerGluGluArgSerArgProAsnSerGlyAspGluLys	126
Db	628	AAAAAGAGCTGTGTGAAGAGCCTGAACACGAGCCCGAAGCTGCAGAGGGTGACGGTGAT	687
QY	127	ArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMetGlyGluGlyLysLysArg	146
Db	688	AAGAAG---GGGATGCAGAGGGCAGCAGCAGCAG-----GAAGGGAGCTGGTC	735
QY	147	ValSerSerGlySerSerGluArgGlySerLysSerProLeuLysArgAlaGln----	164
Db	736	ATTGATGAGCCAGCCAGGAGAAACGAGAAAGGAGCGTTGAAGAGCAGCAGCGGAC	795
QY	165	-----GluInSerProArg-----LysArgGlyArgProProLysAspGluLysAsp	180
Db	796	TTGTCTGGAGGACTCTCTTAAACGTCCCAAGGAGGCAGAAACCCCTGAAGAGAGGAG--	852
QY	181	LeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPhe	200
Db	853	-----AAGGAGGCAGCCACCTTGGAGTTGAGAGGCCCTTCTATTGAGGTGGAA	903
QY	201	LysTrpGlnProThrAlaSerGluProValLysAspAlaAspPro	215
Db	904	ATATGACATCCCTCTGAGCCCGGGCTCTGGCCGGGGGCT	945

DEPT 15

RESULT IS
IIC 00 030 00EA 3

Sequence 2, Application US/09938885A
Patent No. US20020090679A1
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
Goli, Surya K.
Murry, Lynn E.

TITLE OF INVENTION: LUNG GROWTH FACTOR VARIANT

NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/938,885A
FILING DATE: 24-Aug-2001
CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/760,745
FILING DATE: 1996-12-07

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0169 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 869 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: lungast01
CLONE: 876242

SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-09-938-885A-2

Alignment Scores:
Pred. No.: 3,07e-10 Length: 869
Score: 179.50 Matches: 66
Percent Similarity: 43.95% Conservative: 32
Best Local Similarity: 29.60% Mismatches: 100
Query Match: 13.61% Indels: 25
DB: 9 Gaps: 9

US-09-987-755-2 (1-249) x US-09-938-885A-2 (1-869)

Qy 3 AlaValSerLeu-----ArgLeuGlyAspLeuValTrpGlyLysLeuGlyArg 18
Db 35 GCGGTGACATGCGACACACCGCTTCAGCCCGGGGACTTGGTGTTCGCTTAAGATGAGGGC 94
Qy 19 TyrProTrpTrpProGlyLysIleValAsnProProlLysAspLeuLysLysProArgGly 38
Db 95 TACCCTCACTGGCTGCCAGGATCGACACATCGGGATGCGCGTGAAGCCGCCACCC 154
Qy 39 LysLysCysPhePheValLysPhePheGlyThrGluAspHisAlaTrpIleLysValGlu 58
Db 155 AACAAAG---TACCCCATCTTTTCTTTTGGCACACACGAAACAGCCCTCTCGGGCCCAAA 211
Qy 59 GlnLeuLysProTyrHisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArg 78
Db 212 GACCTCTTCCTTACGAGGATCCNAGGAGAGTTTGGCAAGCCCAACACAGAGGAGGG 271

Qy 79 PheGlnGlnAlaValAspAlaValGluGluPheLeuArgArgAlaLysGlyLysGln 98
Db 272 TTCAGCGAGGGGCTGTGGAGATCGAAGC-----AACCTACTGTCAAG 316
Qy 99 ThrSerSerHisAsnSerSerAspAspLysAsnArgAsnSerSerGluGluArgSer 118
Db 317 GCTTCCGGCTATAGTCTCTCCAGAAAAGAGCTGTGTGGAAGAGCCTGAACCCAGAGCCC 376
Qy 119 ArgProAsnSerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValLysLys 138
Db 377 GAAGCTGCAGAGGGTGACGGTGATAAGAG---GGGAATGCAGAGGGCAGCAGCAGCAG 433
Qy 139 AsnMetGlyGluGlyLysLysArgValSerSerGlySerSerGluArgGlySerLysSer 158
Db 434 -----GAAGGGAAGCTGCTCATTGATGAGCCAGCCCAAGAGAGAAAGAAAGGA 484
Qy 159 ProLeuLysArgAlaGln-----GluGlnSerProArg-----LysArgGly 172
Db 485 GCGTTGAAGAGGAGAGCAGGGGACTTGTGTGAGGACTCTCTCTAAACGTCCCAAGAGGCA 544
Qy 173 ArgProProlLysAspGluLysAspLeuThrIleProGluSerSerThrValLysGlyMet 192
Db 545 GAAACCCCTGAAGCAGAGGAG-----AAGGAGGAGCAGCCACTTGGAGGTTGAG 592
Qy 193 MetAlaGlyProMetAlaAlaPheLysTrpGlnProThrAlaSerGluProValLysAsp 212
Db 593 AGGCCCTTCTCTATGAGGTGAAAG---AATAGCACCCCTCTGAGCCCGGCTCTGGC 649
Qy 213 AlaAspPro 215
Db 650 CGGGGGCCT 658

Search completed: December 8, 2003, 12:48:20

Job time : 331 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 8, 2003, 10:34:41 ; Search time 279 Seconds
(without alignments)
2409.176 Million cell updates/sec.

Title: US-09-987-755-2
Perfect score: 1319
Sequence: 1 MAASVRLGLVWVKLGRYP.....VCYQAITKKIKCEDLLLR 249

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPRO_spool/US09987755/runat_08122003_103435_18300/app_query.fasta_1.391
-DB=N Geneseq 19Jun03 -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09987755 @CGN 1 312 @runat_08122003_103435_18300 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N Geneseq 19Jun03.*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT.*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1319	100.0	3202	18	AAT47520	Human hepatoma-der DNA encoding novel
2	1319	100.0	3256	22	AAS33262	CDNA encoding novel
3	1319	100.0	3256	22	AAS34946	CDNA encoding novel
4	1314	99.6	3172	22	AAS34789	Human nucleic acid
5	1291	97.9	1939	25	AA151565	Clone AQ73_3 codin
6	1291	97.9	3316	19	AAV44859	Human cDNA clone A
7	1291	97.9	3316	22	AAF98456	CDNA encoding novel
8	1278	96.9	3707	22	AAS34889	CDNA encoding novel
9	1263	95.8	810	22	AAS35012	EST clone BV27. H
10	1263	95.8	810	22	AAS35012	EST clone BV27. H
C 11	591.5	44.8	658	20	AAV87516	Human immune/haema
12	504	38.2	2850	22	AAK70375	Human immune/haema
13	504	38.2	2850	22	AAK70375	EST clone AQ73. H
C 14	444	33.7	307	20	AAV86389	Human secreted pro
15	341	25.9	445	21	AAV86389	Human secreted pro
16	204.5	15.5	4487	24	ABS70439	Human bone remodel
17	194.5	14.7	3792	23	ABL03572	Drosophila melanog
18	193.5	14.7	1024	21	AAAI3161	Human type II hepa
19	193.5	14.7	1898	22	AAF76854	Human secreted pro
20	193.5	14.7	1973	22	AAH14117	Human cDNA sequenc
21	193.5	14.7	3805	20	AAZ00048	HGF2 gene. Homo
22	193	14.6	1713	23	ABL03573	Drosophila melanog
23	192	14.6	2920	24	ABL35012	Rat cDNA isolated
24	189	14.3	609	21	AAF65004	Human hepatoma-der
25	186.5	14.1	2133	21	AAF18049	Lung cancer associ
26	185.5	14.1	714	18	AAV88420	Mouse hepatoma der
27	185.5	14.1	723	16	AAQ79903	Human hepatoma der
28	185.5	14.1	1386	22	AAV59292	Human hHDGP nucleo
29	185.5	14.1	1583	18	AAV03867	Mouse hepatoma der
30	185.5	14.1	2376	16	AAQ79902	Human hepatoma der
31	185.5	14.1	2376	21	AAA40118	Human HDGF1N cDNA.
32	185.5	14.1	2376	22	ABA83087	Hepatoma-derived g
33	179.5	13.6	869	19	AAV39154	Lung growth factor
34	179	13.6	737	22	AAH05830	Human cDNA clone (
35	177.5	13.5	1919	21	AAH77751	Human cancer assoc
36	166	12.6	1297	22	AAF76856	Human secreted pro
37	166	12.6	2271	21	AAA37108	Human PRO1604 (UNQ
38	166	12.6	2272	22	AAV46097	Human DNA encoding
39	166	12.6	2272	22	AAV46097	DNA encoding prote
40	166	12.6	2272	25	ACA57855	Human PRO1604 cDNA
41	166	12.6	2272	25	ABX98325	Human cDNA encodin
42	166	12.6	2272	25	ABX98325	Novel human secret
43	166	12.6	2272	25	ACA05872	Human secreted/tra
44	166	12.6	2272	25	ABX97916	Human PRO polynucl
45	166	12.6	2272	25	ABX78700	Human PRO polynucl

ALIGNMENTS

RESULT 1
AAT47520
ID AAT47520 standard; cDNA; 3202 BP.
XX
AC AAT47520;
XX
XX 20-MAY-1997 (first entry)
DT
XX
DE Human hepatoma-derived growth factor (HDGF-2) cDNA.
XX
DE Hepatoma-derived growth factor-2; HDGF-2; chemokine; wound healing;
KW vulnery; burn; ulcer; thrombosis; arteriosclerosis; gene therapy;
KW ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers

```

FT CDS 34..783
FT primer_bind /*tag= a
FT complement (34..58)
FT /*tag= b
FT /note= "primer for HDGF-2 recombinant prodn."
XX PN WO9639485-A1.
XX PD 12-DEC-1996.
XX PF 05-JUN-1995; 95WO-US06731.
XX PR 05-JUN-1995; 95WO-US06731.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Kunsch CA, Rosen CA;
XX DR P-PSDB; AAW09404.
XX PT DNA encoding human-derived growth factor polypeptide - useful to
XX PT promote wound healing as result of burns or ulcers
XX PS Claim 1; Fig 1A-C; 54pp; English.
XX CC A cDNA clone (AA747520) codes for human hepatoma-derived growth
XX CC factor-2 (HDGF-2) (AAW09404), a protein that can be used to
XX CC stimulate tissue repair and tissue growth. It was discovered in
XX CC a human umbilical vein endothelial tissue cDNA library, and may
XX CC also be obt'd. from heart, brain and skeletal muscle. HDGF-2
XX CC polynucleotides can be incorporated into vectors for prodn. of
XX CC recombinant HDGF-2 in host (e.g. E. coli, Sf9) cells. They can
XX CC also be used for the gene therapy of diseases related to HDGF-2
XX CC underexpression and for the design diagnostic probes.
XX SQ Sequence 3202 BP; 817 A; 816 C; 819 G; 747 T; 3 other;

Alignment Scores:
Pred. No.: 5,666-105 Length: 3202
Score: 1319.00 Matches: 249
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 18 Gaps: 0

US-09-987-755-2 (1-249) x AAT47520 (1-3202)

QY 1 MetAlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrPro 20
DB 34 ATGGCGGCTGTGAGTCTGGCGCTCGCGACTTGGTGTGGGGAACTCGGCCGATATCT 93
QY 21 ProTrpProGlyLysIleValAsnProProLysAspLeuLysLysProArgGlyLysLys 40
DB 94 CCTTGGCCAGGAAGATTGTTAATCCACCAAGGACTTGAAGAACTCGCGGAAAGAAA 153
QY 41 CysPhePheValLysPhePheGlyThrGluAspHisAlaTrpIleLysValGluClnLeu 60
DB 154 TGCCTCTTTGTGAATTTTGTGGAACAGAGATCATGCTGGATCAAAAGTGGAAACGCTG 213
QY 61 LysProTyrHisAlaHisLysGluMetIleLysIleAsnLysGlyLysArgPheGln 80
DB 214 AAGCATATCATGCTCATAAAGAGAAATGATAAAATTAACAGGGTAAACGATTCCAG 273
QY 81 GlnAlaValAspAlaValGluGluPheLeuArgAlaLysGlyLysAspGlnThrSer 100
DB 274 CAAGCGGTAGATGCTGTGCAAGATTCTCTCAGGAGAGCCAAAGGAAAGAACAGACGTCA 333
QY 101 SerHisAsnSerSerAspAspLysAsnArgArgAsnSerSerGluLysArgSerArgPro 120
DB 334 TCCCAACAATCTTCTGTGACAAAGAAATCGACGTAATTCAGTGGAGAGAGTAGGCCA 393
QY 121 AsnSerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMet 140

394 AACTCAGTGATGAGAAGCGCAAACTTAGCTGTCTGAGGGAAGGTGAAGAACAATG 453
141 GlyCluGlyLysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeu 160
454 GGAGAAGGAAGAAGAGGGGTGTCTTCAGGCTCTTCAGAGAGAGGCTCAAAATCCCTCTG 513
161 LysArgAlaGlnGluInSerProArgLysArgGlyArgProProLysAspGluLysAsp 180
514 AAAAGAGCCCAAGAGCAAGTCCCGGAAGCGGGTGGCCCCCAAGAGGATGAGAAGAT 573
181 LeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPhe 200
574 CTCACCATCCCGAGTCTAGTACCGTGAAGGGGATGATGCCGACCGATGGCCGCGTT 633
201 LysTrpGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisPhe 220
634 AAATGGCAGCCACCCGCAAGCGAGCCTGTTAAAGATGCAGATCTCATTTCCATCTTC 693
221 LeuLeuSerGlnThrGluLysProAlaValCysTyrGlnAlaIleThrLysLysLeuLys 240
694 CTGCTAAGCCAAACAGAGAGCCAGCTGTCTGTACCAGGCATCAGCAAGAAAGTTGAAA 753
241 IleCysGluAspLeuLeuLeuProArg 249
754 ATATGTGAAGACCTCTCTTCTTAGG 780

RESULT 2
AAS33262
ID AAS33262 standard; cDNA; 3256 BP.
AC AAS33262;
XX
XX 04-DEC-2001 (first entry)
XX
XX DNA encoding human secreted protein, Seq ID No 221.
XX Immunomodulatory; human immunodeficiency virus; HIV; anaemia; angina;
XX rheumatoid arthritis; antiarteriosclerotic; cardiant; vascular;
XX cerebroprotective; thrombolytic; antimicrobial; ophthalmological;
XX cycostatic; Alzheimer's disease; Parkinson's disease; human; cancer;
XX multiple sclerosis; cancer; hyperproliferative disorder; infection;
XX Gaucher's disease; neurological disease; cerebrovascular disorder;
XX thrombosis; wound healing; ss.
XX Homo sapiens.
XX
XX WO200155326-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01347.
XX
XX 31-JAN-2000; 2000US-0179065.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-451931/48.
XX
XX P-PSDB; AAU20553.
XX
XX New nucleic acids and polypeptides, useful for diagnosing, preventing
XX or treating medical conditions -
XX
XX Claim 1; SEQ ID No 221; 753pp; English.
XX
XX The invention relates to novel isolated nucleic acid molecules (I)
XX encoding human secreted proteins (II). (I) and (II) are used to prevent,
XX treat or ameliorate a medical condition in e.g. humans, mice, rabbits,
XX goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used
XX in the prevention, treatment and diagnosis of diseases associated with
XX inappropriate expression of secreted proteins. (I) and complementary

```


sequences may also be used as DNA probes in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect and quantitate the presence of similar nucleic acid sequences in samples, and so which patients may be in need of restorative therapy. (ii) may also be used as antigens in the production of antibodies and in assays to identify modulators (agonists and antagonists) of the expression and activity of the secreted proteins. The anti-(ii) antibodies and antigens may also be used to down regulate expression and activity of (ii). The anti-(ii) antibodies may also be used as diagnostic agents for detecting the presence of (ii) in samples (e.g. by enzyme linked immunosorbent assay (ELISA)). The disorders include for example: immune/autoimmune diseases (e.g. HIV (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis and multiple sclerosis), cancers and hyperproliferative disorders (e.g. melanomas, neoplasms of the breast or liver, Sezary syndrome and Gaucher's disease), neurological diseases (e.g. Alzheimer's disease, Parkinson's disease and Charcot-Marie-Tooth disease), cardio-/cerebrovascular disorders (e.g. cardiac arrest, tachycardia, angina and thrombosis), infections caused by bacteria, viruses and fungi and ocular disorders (e.g. corneal infections). (i) and (ii), agonists, antagonists and antibodies can also be used to promote wound healing, maintain organs before transplantation, and support cell culture of primary tissues. AAS33043-AAS33486 represent human secreted protein coding sequences, PCR primers, and related sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification but was obtained in electronic format directly from WIPO at: ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 3256 BP; 828 A; 835 C; 838 G; 753 T; 2 other;

Alignment Scores:

Pred. No.: 5.78e-105 Length: 3256
Score: 1319.00 Matches: 249
Percent Similarity: 100.00% Conservativity: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 22 Gaps: 0

US-09-987-755-2 (1-249) x AAS33262 (1-3256)

QY 1 MetAlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyValTrpPro 20
DB 34 ATGGCGGTGTGAGTCTCGGCTCGGACTGTGTGGGGAACCTCGGCGGATATCCT 93
QY 21 ProTrpProGlyLysIleValAsnProProLysAspLeuLysLysProArgGlyLysLys 40
DB 94 CTTTGGCCAGAAAGATTGTTAATCCACCAAGGACITGAAGAACTCGCGGAAAGAA 153
QY 41 CysPhePheValLysPhePheGlyThrGluAspHisAlaTrpLysValGluGlnLeu 60
DB 154 TGCTTCTTTGTGAATTTTGGAAACAGAGATCATGCTGGATCAAAAGTGGAAACAGCTG 213
QY 61 LysProTrpHisAlaHisLysGluMetIleLysIleAsnLysGlyLysArgPheGln 80
DB 214 AAGCCATATCATGCTCATAAAGAGAAATGATAAAATTAACAAGGGTAAACGATTCCAG 273
QY 81 GlnAlaValAspAlaValGluGluPheLeuArgAlaLysGlyLysAspGlnThrSer 100
DB 274 CAACGGTAGATGTGTGAGAGTTCTTCAGAGAGCCAAAGGGAAGACCAACGCTCA 333
QY 101 SerHisAsnSerSerAspAspLysAsnArgAsnSerSerGluGluArgSerArgPro 120
DB 334 TCCCACAATCTTCTGATGACAAGATTCAGCTTAATTCAGTGAGGAGAGAGTAGGCCA 393
QY 121 AsnSerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMet 140
DB 394 AACTCAGGTGATGAGAGCGCAAACTTAGCCGTGCTGAAGGGAAGGTGAAGAACAATG 453
QY 141 GlyGluGlyLysLysArgValSerSerGlySerGluArgGlySerLysSerProLeu 160
DB 454 CGAAGGGAAGAGAGGTGCTTTCAGGCTTTCAGAGAGAGGCTCCAAATCCCTCTG 513
QY 161 LysArgAlaGlnGluGlnSerProArgLysArgGlyArgProProLysAspGluLysAsp 180

DB 514 AAAAGAGCCCAAGAGCAAAAGTCCCGGAGCGGGTGGCCCCCAAGAGGTGAGAGGAT 573
QY 181 LeuThrIleProGluSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPhe 200
DB 574 CTACCATCCCGGAGTCTAGTACCGTGAAGGGATGATGCGCGGACCGATGGCGCTTT 633
QY 201 LysTrpGlnProThrAlaSerGluProValLysAspAlaAspProHisHisPhe 220
DB 634 AAATGGCAGCAACCGCAGCGAGCCTGTTAAAGATGCAGATCCTCATTTCCATCATTC 693
QY 221 LeuLeuSerGlnThrGluLysProAlaValCysTrpGlnAlaIleThrLysLysLeuLys 240
DB 694 CTCTAAGCCAAACAGAGAGCCAGCTGTCTGTACAGGCAATCACGAAGATTGAAA 753
QY 241 IleCysGluAspLeuLeuLeuProArg 249
DB 754 ATATGTGAAGACCTCTCTTCTTAGG 780
RESULT 3
AAS34946
ID AAS34946 standard; cDNA; 3256 BP.
AC AAS34946;
XX 04-DEC-2001 (first entry)
DE cDNA encoding novel human neoplastic disease associated polypeptide #180.
KW Human; neoplastic disease associated polypeptide; cancer; gene therapy;
KW hyperproliferative disorder; neural disorder; immune system disorder;
KW muscular disorder; reproductive disorder; gastrointestinal disorder;
KW pulmonary disorder; cardiovascular disorder; renal disorder;
KW neuroprotective; cytostatic; anti inflammatory; vasotropic; ss.
OS Homo sapiens.
XX WO200155163-A1.
XX 02-AUG-2001.
XX 17-JAN-2001; 2001WO-US01358.
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.

```
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244674.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.

PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249267.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI WPI; 2001-465558/50.
XX P-PSDB; AAU21747.
XX
XX Novel polypeptides and polynucleotides useful as diagnostic reagents to
PT diagnose diseases or disorders associated with aberrant expression or
PT activity of polypeptides, and for treating cancers, rheumatoid
XX arthritis.
XX
XX Claim 4; SEQ ID No 190; 687pp; English.
PS
XX The present invention relates to the isolation of novel human neoplastic
CC disease associated polypeptides (AAU21568-AAU21851), and cDNA and DNA
CC sequences encoding for these polypeptides. The sequences of the
CC invention are useful in the diagnosis, treatment, prevention and/or
CC prognosis of disorders involving neoplastic disease such as
CC hyperproliferative disorders (e.g. leukemia, bone cancer, bladder
CC cancer, brain stem glioma, adult liver cancer, childhood cerebellar
CC astrocytoma, or Hodgkin's lymphoma). The sequences of the invention may
CC also be useful for treating other disorders such as neural disorders,
CC immune system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
CC and renal disorders. The polynucleotide sequences of the invention are
CC also useful in gene therapy. AAS34767-AAS35050 represent cDNA sequences
CC encoding for the novel human neoplastic disease associated polypeptides
CC of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3256 BP; 828 A; 835 C; 838 G; 753 T; 2 other;

Alignment Scores:
Pred. No.: 5.78e-105 Length: 3256
Score: 1319.00 Matches: 249
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
```

Query Match:	100.00%	Indels:	0
DB:	22	Gaps:	0
US-09-987-755-2 (1-249) x AAS34946 (1-3256)			
QY	1	MetAlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrPro	20
DB	34	ATGGCGGCTGTGAGTCTGGCGCTCGCGACTTGGTGTGGGGAAACTCGCGCATATCTT	93
QY	21	ProTirProGlyLysIleValAsnProProLysAspLeuLysLysProArgGlyLysLys	40
DB	94	CCTTGGCCAGAAAGATTGTATACCAAGAGACTTGAGAAACTCGCGAAGAAA	153
QY	41	CysPhePheValLysPheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeu	60
DB	154	TGCTTCTTTGTGAATTTTGGACAGAAAGATCATGCTCGGATCAAGTGGAAACAGCTG	213
QY	61	LysProTyrHisAlaHisLysGluGluMetIleValIleAsnLysGlyLysArgPheGln	80
DB	214	AAGCCATATCATGCTCATAAAGAGGAAATGATAAAATTAACAAAGGTAAACGATTCCAG	273
QY	81	GlnAlaValAspAlaValGluGluPheLeuArgArgAlaLysGlyLysAspGlnThrSer	100
DB	274	CAGCGGTAGATGCTGTGAGAGTTCTCTCAGAGAGCCAAAGGAAAGACAGAGCTCA	333
QY	101	SerHisAsnSerSerAspLysAsnArgArgAsnSerSerGluGluArgSerArgPro	120
DB	334	TCCCACAATTTCTGTATGACAAAGATCGACGTAATTTCCAGTCAGGAGAGAGTAGGCCA	393
QY	121	AsnSerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMet	140
DB	394	AACTCAGGTGATGAGAGGCCAAATTTAGCTGTCTGAGGGAAGGTGAAGAAAGACATG	453
QY	141	GlyGluGlyLysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeu	160
DB	454	GGAGAGGAAGAAGAGGGGTCTTTCAGGCTCTTCAGAGAGAGCTCCAAATCCCCTCTG	513
QY	161	LysArgAlaGlnGluGlnSerProArgLysArgGlyArgProProLysAspGluLysAsp	180
DB	514	AAAAGAGCCCAAGAGCAAGTCCCGAGCGGGTGGCCGCCCAAGGATGAGAAGGAT	573
QY	181	LeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPhe	200
DB	574	CTCACCATCCGGAGTCTAGTCCGTGAAGGGATGATGCCCGGACCGCATGGCGGCTTT	633
QY	201	LysTrpGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisPhe	220
DB	634	AAATGGCAGCCCAACCGCAAGCGAGCTGTTAAGATGCAGATCTCTCATTTCCATCTTC	693
QY	221	LeuLeuSerGlnThrGluLysProAlaValCysTyrGlnAlaIleThrLysLysLeuLys	240
DB	694	CTGCTAAGCCAAACAGAGAGCCAGCTGTCTGTTACCAGCAATCAGAAAGATTGAAA	753
QY	241	IleCysGluAspLeuLeuProArg	249
DB	754	ATATGTGAGACCTCTTCTTCCTAGG	780
RESULT 4			
AAS34789			
ID	AAS34789 standard; cDNA; 3172 BP.		
XX			
AC	AAS34789;		
XX			
DT	04-DEC-2001 (first entry)		
XX			
DE	cDNA encoding novel human neoplastic disease associated polypeptide #23.		
XX			
KW	Human; neoplastic disease associated polypeptide; cancer; gene therapy;		
KW	hyperproliferative disorder; neural disorder; immune system disorder;		
KW	muscular disorder; reproductive disorder; gastrointestinal disorder;		
KW	pulmonary disorder; cardiovascular disorder; renal disorder;		
KW	neuroprotective; cytostatic; anti inflammatory; vasotropic; ss.		
XX			

OS	Homo sapiens.
XX	
PN	WO200155163-A1.
XX	
PD	02-AUG-2001.
XX	
PF	17-JAN-2001; 2001WO-US01358.
XX	
PR	31-JAN-2000; 2000US-0179065.
PR	04-FEB-2000; 2000US-0180628.
PR	24-FEB-2000; 2000US-0184664.
PR	02-MAR-2000; 2000US-0186350.
PR	16-MAR-2000; 2000US-0189874.
PR	17-MAR-2000; 2000US-0190076.
PR	18-APR-2000; 2000US-0198123.
PR	19-MAY-2000; 2000US-0205515.
PR	07-JUN-2000; 2000US-0209467.
PR	28-JUN-2000; 2000US-0214886.
PR	30-JUN-2000; 2000US-0215135.
PR	07-JUL-2000; 2000US-0216647.
PR	11-JUL-2000; 2000US-0217487.
PR	11-JUL-2000; 2000US-0217496.
PR	26-JUL-2000; 2000US-0220963.
PR	26-JUL-2000; 2000US-0220964.
PR	14-AUG-2000; 2000US-0224518.
PR	14-AUG-2000; 2000US-0224519.
PR	14-AUG-2000; 2000US-0225213.
PR	14-AUG-2000; 2000US-0225214.
PR	14-AUG-2000; 2000US-0225266.
PR	14-AUG-2000; 2000US-0225267.
PR	14-AUG-2000; 2000US-0225268.
PR	14-AUG-2000; 2000US-0225270.
PR	14-AUG-2000; 2000US-0225447.
PR	14-AUG-2000; 2000US-0225757.
PR	14-AUG-2000; 2000US-0225758.
PR	14-AUG-2000; 2000US-0225759.
PR	18-AUG-2000; 2000US-0226279.
PR	22-AUG-2000; 2000US-0226681.
PR	22-AUG-2000; 2000US-0226868.
PR	22-AUG-2000; 2000US-0227182.
PR	23-AUG-2000; 2000US-0227009.
PR	30-AUG-2000; 2000US-0228924.
PR	01-SEP-2000; 2000US-0229287.
PR	01-SEP-2000; 2000US-0229343.
PR	01-SEP-2000; 2000US-0229344.
PR	01-SEP-2000; 2000US-0229345.
PR	05-SEP-2000; 2000US-0229509.
PR	05-SEP-2000; 2000US-0229513.
PR	06-SEP-2000; 2000US-0230437.
PR	06-SEP-2000; 2000US-0230438.
PR	08-SEP-2000; 2000US-0231242.
PR	08-SEP-2000; 2000US-0231243.
PR	08-SEP-2000; 2000US-0231244.
PR	08-SEP-2000; 2000US-0231413.
PR	08-SEP-2000; 2000US-0231414.
PR	08-SEP-2000; 2000US-0232080.
PR	08-SEP-2000; 2000US-0232081.
PR	12-SEP-2000; 2000US-0231968.
PR	14-SEP-2000; 2000US-0232397.
PR	14-SEP-2000; 2000US-0232398.
PR	14-SEP-2000; 2000US-0232399.
PR	14-SEP-2000; 2000US-0232400.
PR	14-SEP-2000; 2000US-0232401.
PR	14-SEP-2000; 2000US-0233063.
PR	14-SEP-2000; 2000US-0233064.
PR	14-SEP-2000; 2000US-0233065.
PR	21-SEP-2000; 2000US-0234223.
PR	21-SEP-2000; 2000US-0234274.
PR	25-SEP-2000; 2000US-0234997.
PR	25-SEP-2000; 2000US-0234998.
PR	26-SEP-2000; 2000US-0235484.


```

Db 429 GAAGGAAAGAGAGGGTGTCTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTGAAA 488
Qy 162 ArgAlaGlnGluGlnSerProArgLysArgGlyValArgProProLysAspGluLysAspLeu 181
Db 489 AGAGCCCAAGAGCAAAATCCCGAGAGGGGTCCGCCCAAGAGGATGAGAGGATCTC 548
Qy 182 ThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPheLys 201
Db 549 ACCATCCCGAGTCTAGTACCGTGAAGGGGATGATGCCGACCGATGGCGGCTTTAAA 608
Qy 202 TrpGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisPheLeu 221
Db 609 TGGCAGCCCAACCGCAAGCAGCTGTATAAGATCAGATCCTCATTTCCATCATTTCCCTG 668
Qy 222 LeuSerGlnThrGluLysProAlaValCysTyrGlnAlaIleThrLysLysLeuLysIle 241
Db 669 CTAGCCCAACAGAGAGCAGCTGTCTGTACAGGCAATCAGCAAGAGTTGAAATA 728
Qy 242 CysGluAspLeuLeuLeuProArg 249
Db 729 TGTGAGACCTCTTCTTCCTAGG 752

RESULT 5
AAL51565
ID AAL51565 standard; DNA; 1939 BP.
XX
AC AAL51565;
XX
DT 10-APR-2003 (first entry)
XX
DE Human nucleic acid-associated protein coding sequence - SEQ ID NO 48.
XX
KW Human; gene; ds; nucleic acid-associated protein; NAAP; arteriosclerosis;
KW cell proliferative disorder; atherosclerosis; cirrhosis; hepatitis; AIDS;
KW cancer; developmental disorder; renal tubular acidosis; anaemia; asthma;
KW mental retardation; neurological disorder; Alzheimer's disease; epilepsy;
KW Parkinson's disease; autoimmune disorder; inflammatory disorder; allergy;
KW Crohn's disease; transgenic animal; animal model.
XX
OS Homo sapiens.
XX
PN WO2003000864-A2.
XX
PD 03-JAN-2003.
XX
PF 20-JUN-2002; 2002WO-US21179.
XX
PR 22-JUN-2001; 2001US-300518P.
PR 29-JUN-2001; 2001US-301787P.
PR 29-JUN-2001; 2001US-301792P.
PR 29-JUN-2001; 2001US-301892P.
PR 29-JUN-2001; 2001US-301893P.
PR 06-JUL-2001; 2001US-303405P.
PR 06-JUL-2001; 2001US-303442P.
PR 15-MAR-2002; 2002US-364438P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Gandhi AR, Swarnakar A, Hafalia AJA, Warren BA, Emerling BM;
PI Arvizu CS, Ison CH, Honchell CD, Lee EA, Yue H, Forsythe IJ;
PI Ramkumar J, Griffin JA, Yang J, Sanjanwala MM, Baughn NR;
PI Borowsky ML, Yao MG, Walia NK, Bandman O, Lal PG, Becha SD;
PI Lee SY, Richardson TW, Elliott VS, Luo W, Tang YT, Zebarjadian Y;
PI Lu Y;
XX
DR WPI; 2003-201420/19.
DR P-PSDB; AAO16415.
XX
PT New nucleic acid-associated proteins and polynucleotides, useful for
PT diagnosing, treating or preventing cell proliferative (e.g. cancer),
PT neurological (e.g. epilepsy or Parkinson's disease), or autoimmune
PT disorders (e.g. AIDS) -

```

```

XX Claim 12; Page 285-286; 312pp; English.
XX
CC The invention comprises the amino acid and coding sequences of human
CC nucleic acid-associated proteins (NAAP). The DNA and protein sequences of
CC the invention are useful for diagnosing, treating or preventing disorders
CC associated with aberrant expression of NAAP, such as: cell proliferative
CC disorders (e.g. arteriosclerosis, atherosclerosis, cirrhosis, hepatitis
CC or cancer); developmental disorders (e.g. renal tubular acidosis, anaemia
CC or mental retardation); neurological disorders (e.g. Alzheimer's disease,
CC Parkinson's disease or epilepsy); and autoimmune/inflammatory disorders
CC (e.g. AIDS, allergies, asthma or Crohn's disease). The DNA sequences of
CC the invention are useful for creating transgenic animals to model human
CC disease. The present DNA sequence encodes a human nucleic acid-associated
CC protein of the invention.
XX
SQ Sequence 1939 BP; 504 A; 498 C; 548 G; 389 T; 0 other;
XX
Alignment Scores:
Pred. No.: 8.41e-103 Length: 1939
Score: 1291.00 Matches: 243
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.59% Mismatches: 0
Query Match: 97.88% Indels: 0
DB: 25 Gaps: 0

US-09-987-755-2 (1-249) x AAL51565 (1-1939)
Qy 1 MetAlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrPro 20
Db 9 ATGGCGGCTGTGAGTCTGCGGCTCGCGGACTTGGTGGGGGAAACTCGGCCGATATCCT 68
Qy 21 ProTrpProGlyLysIleValAsnProProLysAspLeuLysLysProArgGlyLysLys 40
Db 69 CTTGGCCAGAAAGATTGTATTACACCAAGAGCTTGAAGAACTTCGGGAAAGAA 128
Qy 41 CysPhePheValLysPheGlyThrGluAspHisAlaTrpLysValGluGlnLeu 60
Db 129 TGCTTCTTTGTAATTTTGGACACAGAGATCATGCTGGATCAAAAGTGGAAACAGCTG 188
Qy 61 LysProTyrHisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGln 80
Db 189 ARGCCATATCATGCTCATAAAGAGGAAATGATAAAATTAAACAGGGTAAACCGATTCCAG 248
Qy 81 GlnAlaValAspAlaValGluGluPheLeuArgArgAlaLysGlyLysAspGlnThrSer 100
Db 249 CAAGCGGTAGATGCTGTGAGAGAGTTCCTCAGAGAGCCCAAGGAAAGACGACGCTCA 308
Qy 101 SerHisAsnSerSerAspLysAsnArgArgAsnSerSerGluGluArgSerArgPro 120
Db 309 TCCACAAATCTTCTGATGACAAGATCGACGTAATTTCCAGTGAGGAGAGAGTAGGCCA 368
Qy 121 AsnSerGlyAspGluLysArgLysLeuSerLeuSerSerGluGlyLysValLysLysAsnMet 140
Db 369 AACTCAGGTGATGAGAGCGCAAACTTATAGCTGTCTGAAGGGGAAAGGTGAAGAAACATG 428
Qy 141 GlyGluGlyLysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeu 160
Db 429 GGAAGAGAAAGAGAGGGGTGCTTTCAGGCTCTTCAGAGAGAGGCTCAAAATCCCTCTG 488
Qy 161 LysArgAlaGlnGluGlnSerProArgLysArgGlyArgProProLysAspGluLysAsp 180
Db 489 AAAAGAGCCCAAGAGCAAAAGTCCCGGAGCGGGTCCGCCCAAGAGGATGAGAGGAT 548
Qy 181 LeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPhe 200
Db 549 CTCACCATCCCGAGTCTAGTACCGTGAAGGGGATGATGGCCGACGATGGCGCGCTTT 608
Qy 201 LysTrpGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisPhe 220
Db 609 AAATGGCAGCCCAACCGCAAGCAGCTGTATAAGATCAGATCCTCATTTCCATCATTC 668
Qy 221 LeuLeuSerGlnThrGluLysProAlaValCysTyrGlnAlaIleThrLysLysLeuLys 240

```

```

Db 669 CTGCTAAGCCAAACAGAGAGCCAGCTGTCTGTACCGGCAATCACGAAGAAGTTGAAA 728
Qy 241 IleCysGluAasp 244
Db 729 ATATGTGAAGAG 740
RESULT 6
AAV44859
ID AAV44859 standard; cDNA; 3316 BP.
XX AAV44859;
XX
DT 21-OCT-1998 (first entry)
XX
DE Clone AQ73_3 coding sequence.
XX
KW Secreted protein; nutritional source; cell proliferation activity;
KW cell differentiation activity; immune stimulant; tissue growth activator;
KW haematopoiesis regulator; anti-inflammatory; tumour invasion suppressor;
KW tumour inhibitor; clone AQ73_3; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 7..1650
FT /tag= a
XX
XX WO9825962-A2.
XX
XX 18-JUN-1998.
PD
PF 12-DEC-1997; 97WO-US23224.
XX
PR 11-DEC-1997; 97US-0989232.
PR 13-DEC-1996; 96US-0766263.
XX
XX (GENY ) GENETICS INST INC.
XX
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
PI Racie LA, Spaulding V, Treacy M;
XX
DR WPI; 1998-362424/31.
DR P-PSDB; AAW69240.
XX
XX New isolated polynucleotides - obtained from human adult testis,
PT human adult ovary, human adult brain and human adult heart cDNA
PT libraries
XX
PS Claim 14; Page 63-65; 108pp; English.
XX
CC This sequence represents a polynucleotide of the invention, and encodes a
CC secreted protein. It was isolated from a human adult ovary cDNA library,
CC and is designated clone AQ73_3. The DNA sequences and encoded
CC polypeptides can be used as nutritional sources or supplements, or may
CC exhibit e.g. cytokine and cell proliferation/differentiation activity,
CC immune stimulating or suppressing activity, haematopoiesis regulating
CC activity, receptor/ligand activity, anti-inflammatory activity,
CC activin/inhibin activity, chemostatic/chemokinetic activity,
CC cadherin/tumour invasion suppressor activity, tissue growth activity,
CC tumour inhibition activity or other activities.
XX
SQ Sequence 3316 BP; 875 A; 813 C; 887 G; 741 T; 0 other;

Alignment Scores:
Pred. No.: 1.6e-102 Length: 3316
Score: 1291.00 Matches: 243
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.59% Mismatches: 0
Query Match: 97.88% Indels: 0
DB: 19 Gaps: 0

US-09-987-755-2 (1-249) x AAV44859 (1-3316)

```

```

Qy 1 MetAlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrPro 20
Db 7 ATGGCGGCTGTGAGCTCTGGGCTCGGCGAGCTTGTGTGGGGAAACTCGGCCGATATCCT 66
Qy 21 ProTrpProGlyLysIleValAsnProProLysAspLeuLysLysProArgGlyLysLys 40
Db 67 CCTTGGCCAGGAAAGATTGTTAATCCACCAAGGACTTGAAGAAACCTCGCGGAAAGAAA 126
Qy 41 CysPhePheValLysPheGlyThrGluAspHisAlaTrpIleLeuValGluGlnLeu 60
Db 127 TGTCTTCTTTTGAAATTTTTTGGAAACAGAGATCATGCTCGATCAAGTGGAAACAGCTG 186
Qy 61 LysProTyrHisAlaHisLysLysGluGluMetIleLysIleAsnLysGlyLysArgPheGln 80
Db 187 AAGCCATATCATGCTCATAAAGAGGAAATGATAAAATTAACAAGGGTAAACGATTCCAG 246
Qy 81 GlnAlaValAspAlaValGluGluPheLeuArgAlaLysGlyLysAspGlnThrSer 100
Db 247 CAAGCGGTAGATGCTGTCTGAAGAGTTCTCTCAGGAGAGCCAAAGGAAAGACCAACGTC 306
Qy 101 SerHisAsnSerSerAspAspLysAsnArgArgAsnSerSerGluGluArgSerArgPro 120
Db 307 TCCCACAATTTCTCTGATGACAGATCGAGTAAATTCAGTGAGGAGAGAGTAGGCCA 366
Qy 121 AsnSerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMet 140
Db 367 AACTCAGGTGATGAGAGGCGCAACTTAGCCCTGCTCGAAGGGAAGGTGAAGAACAATG 426
Qy 141 GlyGluGlyLysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeu 160
Db 427 GGAGAAGGAAAGAGAGGGGTGCTCTTCAGGGCTTTTCAGAGAGAGGCTCCAAATCCCTCTG 486
Qy 161 LysArgAlaGlnGluGlnSerProArgLysArgGlyValArgProProLysAspGluLysAsp 180
Db 487 AAAAGAGCCCAAGAGCAAAAGTCCCGGAAGCGGGGTGGCCCCCAAGAGGTAGAAGANT 546
Qy 181 LeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPhe 200
Db 547 CTCACCATCCCGAGTCTAGTACCGTGAAGGGGATGATGCCGAGCCGATGCCCGCTTT 606
Qy 201 LysTrpGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisPhe 220
Db 607 AAATGGCAGCCACCGCAAGCGAGCTGTAAAGATGCAGATCCTCATTTCCATCATTTTC 666
Qy 221 LeuLeuSerGlnThrGluLysProAlaValCysTyrGlnAlaIleThrLysLysLysLys 240
Db 667 CTGCTAAGCCCAACAGAGAGCCAGCTGTCTGTACCAGGCATCACGAAGAAGTTGAAA 726
Qy 241 IleCysGluAasp 244
Db 727 ATATGTGAAGAG 738
RESULT 7
AAV98456
ID AAV98456 standard; cDNA; 3316 BP.
XX
AC AAV98456;
XX
DT 07-JUN-2001 (first entry)
XX
DE Human cDNA clone AQ73_3 sequence SEQ ID 136.
XX
KW Human; secreted protein; nutrient; cytokine modulator; proliferation;
KW differentiation; immune system modulator; tissue growth; chemotactic;
KW haemostatic; thrombolytic; anti-inflammatory; tumour inhibition; ss;
KW haematopoiesis.
XX
OS Homo sapiens.
XX
PN WO200119988-A1.
XX
PD 22-MAR-2001.

```

```

XX 14-SEP-2000; 2000WO-US25135.
XX PF
XX 17-SEP-1999; 99US-0398829.
XX PR
XX (GEMY ) GENETICS INST INC.
XX PA
XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
XX PI Merberg D, Treacy M, Bowman MR, Spaulding V, Agostino MJ;
XX PR
XX WPI: 2001-244801/25.
XX DR P-PSDB; AAB90720.
XX DR
XX Isolated nucleic acids encoding polypeptides, useful for modulating
XX PT e.g. cytokine and cell proliferation/differentiation activity, the
XX PT immune system and hematopoiesis regulating activity -
XX XX
XX Disclosure; Page 467-468; 557pp; English.
XX XX
XX Human cDNA clones represented in AAF98374 - AAF98489 encode secreted
XX CC proteins AAB90667 - AAB90750. The cDNA clones are isolated from various
XX CC tissue types, and may be used in the prevention, treatment and diagnosis
XX CC of diseases associated with inappropriate protein expression. The
XX CC polypeptides and nucleic acids may be used as nutrients or to modulate
XX CC cytokine and cell proliferation/differentiation activity and may also be
XX CC involved in modulation of the immune system. The cDNA sequences,
XX CC proteins, their agonists and/or antagonists exhibit haematopoiesis
XX CC regulating activity; tissue growth activity; activin/inhibin activity;
XX CC chemotactic/chemokinetic activity; haemostatic and thrombolytic
XX CC activity; receptor/ligand activity; anti-inflammatory activity;
XX CC haematopoiesis activity; cadherin/tumour suppressor activity; and/or
XX CC tumour inhibition activity. Included in the invention are probes
XX CC represented in AAF98490 - AAF98572 which are specific for the cDNA clones
XX CC encoding the secreted proteins.
XX XX
XX Sequence 3316 BP; 875 A; 813 C; 887 G; 741 T; 0 other;
XX SQ

```

Alignment Scores:

Pred. No.:	1-6e-102	Length:	3316
Score:	1291.00	Matches:	243
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	99.59%	Mismatches:	0
Query Match:	97.88%	Indels:	0
DB:	22	Gaps:	0

US-09-987-755-2 (1-249) x AAF98456 (1-3316)

```

Qy 1 MetAlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrPro 20
Db 7 ATGGCGGCTGTGAGTCTGCGGCTCGCGACTTGGTGGGGAACTCGGCGGATATCCT 66
Qy 21 ProTrpProGlyLysIleValAsnProProLysAspLeuLysProArgGlyLysLys 40
Db 67 CCTTGGCCAGAAAGATTTGTAATCCACCAAGACTTGAAGAACTCGCGGAAAGAA 126
Qy 41 CysPhePheValLysPheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeu 60
Db 127 TGCCTTCTTTGTAATTTTGGAAACAGAGATCATGCTGGATCAAAAGTGGACAGCTG 186
Qy 61 LysProTyrHisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGln 80
Db 187 AAGCCATATCATGCTCATAAAGAGAAATGATAAAATTAACAAAGGTTAAACGATTCAG 246
Qy 81 GlnAlaValAspAlaValGluPheLeuArgArgAlaLysGlyLysAspGlnThrSer 100
Db 247 CAACCGGTAGATGCTGTGCAAGAGTTCTCTCAGGAGAGCCAAAGGAAAGACCAAGCTCA 306
Qy 101 SerHisAsnSerSerAspLysAsnArgArgAsnSerSerGluGluArgSerArgPro 120
Db 307 TCCCACAATCTCTTGATGACAGAAATCGAGTAAATTCAGTGGAGAGAAAGTAGGCCA 366
Qy 121 AsnSerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMet 140

```

367 AACTCAGGTGATGAGAAGCGCAAACTTAGCTGCTCTGAAGGGAGGTGAAGAAACATG 426

141 GlyGluGlyLysLysArgValSerSerGlySerGluArgGlySerLysSerProLeu 160

427 GGNAGAGGAAAGNAGAGGGGTGCTTCAGGCTCTTCAGAGAGAGCTCCAAATCCCTCTG 486

161 LysArgAlaGlnGlnSerProArgLysArgGlyArgProProLysAspGluLysAsp 180

487 AAAGAGCCCAAGAGCAAAAGTCCCGAAGCGGGTCCGCCCCCAAGAGATGAGAAGGAT 546

181 LeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPhe 200

547 CTCACCATCCCGAGTCTAGTACCGTGAAGGGGATGATGCCCGGACCGATGGCGCGTTT 606

201 LysTrpGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisPhe 220

607 AANTGCAGCCCAACCCGAGGAGCTGTTAAGATGCAGATCTCATTTCCATCATTTTC 666

221 LeuLeuSerGlnThrGluLysProAlaValCysTyrGlnAlaIleThrLysLysLeuLys 240

667 CTGCTAAGCCAAACAGAGAACCGAGCTGCTGTACGAGCAATCAGAAAGATTGAAA 726

241 IleCysGluAsp 244

727 ATATGTGAAGAG 738

RESULT 8

AAS34889

ID AAS34889 standard; cDNA; 3707 BP.

XX AAS34889;

AC AAS34889;

XX

DT 04-DEC-2001 (first entry)

XX

DE cDNA encoding novel human neoplastic disease associated polypeptide #123.

XX

Human; neoplastic disease associated polypeptide; cancer; gene therapy;

KW hyperproliferative disorder; neural disorder; immune system disorder;

KW muscular disorder; reproductive disorder; gastrointestinal disorder;

KW pulmonary disorder; cardiovascular disorder; renal disorder;

XX neuroprotective; cytostatic; anti inflammatory; vasotropic; ss.

OS Homo sapiens.

XX

PN WO200155163-A1.

XX

PD 02-AUG-2001.

XX

PF 17-JAN-2001; 2001WO-US01358.

XX

PR 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249284.
PR 17-NOV-2000; 2000US-0249285.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 05-JAN-2001; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465558/50.
XX P-PSDB; AAU21690.
XX Novel polypeptides and polynucleotides useful as diagnostic reagents to
XX diagnose diseases or disorders associated with aberrant expression or
XX activity of polypeptides, and for treating cancers, rheumatoid
XX arthritis -
XX Claim 4; SEQ ID No 133; 687pp; English.
XX The present invention relates to the isolation of novel human neoplastic
XX disease associated polypeptides (AAU21568-AAU21851), and cDNA and DNA
XX sequences encoding for these polypeptides. The sequences of the
XX invention are useful in the diagnosis, treatment, prevention and/or
XX prognosis of disorders involving neoplastic disease such as
XX hyperproliferative disorders (e.g. leukaemia, bone cancer, bladder
XX cancer, brain stem glioma, adult liver cancer, childhood cerebellar
XX astrocytoma, or Hodgkin's lymphoma). The sequences of the invention may
XX also be useful for treating other disorders such as neural disorders,
XX immune system disorders, muscular disorders, reproductive disorders,
XX gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
XX and renal disorders. The polynucleotide sequences of the invention are
XX also useful in gene therapy. AAS34767-AAS35050 represent cDNA sequences
XX encoding for the novel human neoplastic disease associated polypeptides
XX of the invention.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.


```
DB: 22 Gaps: 0
US-09-987-755-2 (1-249) x AAS33224 (1-810)
QY 4 ValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyProProTrpPro 23
Db 2 GTGAGTCTGCNGCTCGGCGACTGTGTGGGGAAACTCGNGCATATCCTCTTGGCCA 61
QY 24 GlyLysIleValAsnProProLysAspLeuLysLysProArgGlyLysLysCysPhe 43
Db 62 GGAAGATTGTAAATCCACCAAGACCTTGAGAAACCTCGCGGAAAGAAATGCTTCTTT 121
QY 44 ValLysPheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeuLysProTyr 63
Db 122 GTGAAATTTTGGAAACAGACATCATGCTGATCAAGTGAAGTGAACAGTGAAGCCATAT 181
QY 64 HisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGlnAlaVal 83
Db 182 CATGCTCATAAAGAGGAAATGATAAAATTAACAAGGGTAAACGATTCCAGCAAGCGTA 241
QY 84 AspAlaValGluGluPheLeuArgAlaLysGlyLysAspGlnThrSerSerHisAsn 103
Db 242 GATGCTGTGGAAGATTCTCAGAGAGCCAAAGGGAAGACCAAGCATCTATCCCAAT 301
QY 104 SerSerAspAspLysAsnArgAsnSerSerGluArgSerArgProAsnSerGly 123
Db 302 TCTTCTGATGACAAGATCGACGTAAATCCAGTGAGGAGAGAGTAGGCCAAACTCAGGT 361
QY 124 AspGluLysArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMetGlyGluGly 143
Db 362 GATGAGAAGCGCAACTTAGCTGTCTGAAGGGAAGGTGAAGAAACATGGGAGAAGGA 421
QY 144 LysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeuLysArgAla 163
Db 422 AAGAAGAGGGTGTCTTACGCTCTTCAGAGAGAGCTCAAAATCCCTCTGAAGAAGGCC 481
QY 164 GlnGluGlnSerProArgLysArgGlyArgProProLysAspGluLysAspLeuThrIle 183
Db 482 CAAGAGCAAAAGTYCCCGAAGCGGGTTCGGCCCCCAAGAGGTAGAGAGATCTCACCATC 541
QY 184 ProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPheLysTrpGln 203
Db 542 CGGAGTCTAGTACCGTGAAGGGGATGATGCCCGGACCGCATGGCCGCTTTAAATGGCAG 601
QY 204 ProThrAlaSerGluProValLysAspAlaAspProHisPheHisPheLeuLeuSer 223
Db 602 CCAACCGAAGCGAGCTGTAAAGATGCAATCTCATTTCCATCATTTCTGCTAAGC 661
QY 224 GlnThrGluLysProAlaValCysTyrGlnAlaIleThrLysLysLeuLysIleCysGlu 243
Db 662 CAACAGAGAGCCAGCTGTCTGTATACAGGCAATCAGCAAGAGTTCGAAATATGTGAA 721
QY 244 Asp 244
Db 722 GAG 724
RESULT 10
AAS35012
ID AAS35012 standard; cDNA; 810 BP.
XX
AC AAS35012;
XX
DT 04-DEC-2001 (first entry)
XX
DE cDNA encoding novel human neoplastic disease associated polypeptide #246.
XX
KW Human; neoplastic disease associated polypeptide; cancer; gene therapy;
KW hyperproliferative disorder; neural disorder; immune system disorder;
KW muscular disorder; reproductive disorder; gastrointestinal disorder;
KW pulmonary disorder; cardiovascular disorder; renal disorder;
KW neuroprotective; cytostatic; anti inflammatory; vasotropic; ss.
OS Homo sapiens.
```

```
XX WO200155163-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01358.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
```

PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249284.
PR 17-NOV-2000; 2000US-0249285.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
(HUMA-) HUMAN GENOME SCI INC.
XX
XX

PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465558/50.
DR P-PSDB; AAU21813.
XX Novel polypeptides and polynucleotides useful as diagnostic reagents to
PT diagnose diseases or disorders associated with aberrant expression or
PT activity of polypeptides, and for treating cancers, rheumatoid
PT arthritis
XX
PS Claim 4; SEQ ID No 256; 687pp; English.
XX
CC The present invention relates to the isolation of novel human neoplastic
CC disease associated polypeptides (AAU21568-AAU21851), and cDNA and DNA
CC sequences encoding for these polypeptides. The sequences of the
CC invention are useful in the diagnosis, treatment, prevention and/or
CC prognosis of disorders involving neoplastic disease such as
CC hyperproliferative disorders (e.g. leukaemia, bone cancer, bladder
CC cancer, brain stem glioma, adult liver cancer, childhood cerebellar
CC astrocytoma, or Hodgkin's lymphoma). The sequences of the invention may
CC also be useful for treating other disorders such as neural disorders,
CC immune system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
CC and renal disorders. The polynucleotide sequences of the invention are
CC also useful in gene therapy. AAS34767-AAS35050 represent cDNA sequences
CC encoding for the novel human neoplastic disease associated polypeptides
CC of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 810 BP; 259 A; 178 C; 218 G; 152 T; 3 other;

Alignment Scores:
Pred. No.: 7,99e-101 Length: 810
Score: 1263.00 Matches: 238
Percent Similarity: 99.17% Conservative: 1
Best Local Similarity: 98.76% Mismatches: 2
Query Match: 95.75% Indels: 0
DB: 22 Gaps: 0

US-09-987-755-2 (1-249) x AAS35012 (1-810)

Qy 4 ValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrProTrpPro 23
Db 2 GTGAGTCTGCNGCTCGCGGACTTGGTGTGGGGAACTCGGCGATATCTCTTGSCCA 61

Qy 24 GlyLysIleValAsnProLysAspLeuLysLysProArgGlyLysLysCysPhePhe 43
Db 62 GGAAGATTGTTAATCCACCAAGGACTTGAAGAACTCCGCGAAAGAAATGCTTCTTT 121

Qy 44 ValLysPhePheGlyThrGluAspHisAlaTrpIleLysValGluGlnLysProTyr 63
Db 122 GTGAAATTTTTTGAACAGAAAGATCATGCTGTGATCAAGTGAAGACAGTGAAGCCATAT 181

Qy 64 HisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGlnAlaVal 83
Db 182 CATGCTCATTAAGAGGAAATGATATAAAATTAACAGGGTAAACGATTCCAGCAAGCGTA 241

Qy 84 AspAlaValGluGluPheLeuArgAlaLysGlyLysAspGlnThrSerSerHisAsn 103
Db 242 GATGCTGTGCAAGAGTTCCTCAGAGAGCCAAAGGAAAGACAGACGTCATCCCAAT 301

Qy 104 SerSerAspLysAsnArgArgAsnSerSerGluLysArgArgProAsnSerGly 123
Db 302 TCTTCTGATCAAGAAATCGACGTAATTCAGTGAGAGAGAGTAGGCCAAATCTAGGT 361

Qy 124 AspGluLysArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMetGlyGluGly 143
Db 362 GATGAGAAAGCGCAAACTTAGCTGTCTGAAGGGAGAGTGAAGAACATCGGAGAGGA 421

Qy 144 LysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeuLysArgAla 163
|||

Db 422 AAGAGAGGGTCTTTCAGGCTCTTCAGAGAGAGGCTCAAATCCCTCTCTGAAAGAGCC 481

Qy 164 GlnGluGlnSerProArgLysArgGlyArgProProLysAspGluLysAspLeuThrIle 183

Db 482 CAAAGAGCAAGTYCCCGAAGCGGGTCGGCCCCCAAGAGGATGAGAAGATCTCACCATC 541

Qy 184 ProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPheLysTrpGln 203

Db 542 CCGAGTCTAGTACCGTGAAGGGATGATGCGCGACCGATGCGCGCGTCTTAAATGGCAG 601

Qy 204 ProThrAlaSerGluProValLysAspAlaAspProHisPheHisPheLeuLeuSer 223

Db 602 CCAACCGCAAGCGAGCCTGTTAAAGATGCAGATCTCATTTCCATCTTCTGCTAAGC 661

Qy 224 GlnThrGluLysProAlaValCysTrpGlnAlaIleThrLysLysLeuLysIleCysGlu 243

Db 662 CAAACAGAGAGCCAGCTGCTGTTTACCAGGCAATCAGAGAGATTGAAATATGTGA 721

Qy 244 Asp 244

Db 722 GAG 724

RESULT 11

AAV87516/c

ID AAV87516 standard; cDNA; 658 BP.

XX

AC AAV87516;

DT 27-APR-1999 (first entry)

XX

DE EST clone BV27.

XX

KW Expressed sequence tag; secreted protein; haematopoiesis regulator;

KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;

KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;

XX receptor; ligand; anti-inflammatory; tumour inhibitor; ds.

OS Homo sapiens.

XX

PN WO9845435-A2.

XX

PD 15-OCT-1998.

XX

PF 10-APR-1998; 98WO-US06954.

XX

PR 10-APR-1997; 97US-0835913.

XX

PA (GENY) GENETICS INST INC.

XX

PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;

PI Racie LA, Spaulding V, Treacy M;

XX

DR WPI; 1999-070076/06.

XX

XX New polynucleotides encoding human secreted proteins - derived from

PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,

PT ovary, pituitary, retina and colon cDNA libraries

XX

PS Claim 1; Page 589; 633pp; English.

XX

This sequence represents an expressed sequence tag (EST), and is a polynucleotide of the invention. The polynucleotides of the invention are all secreted EST sequences isolated from a variety of human tissue sources. The EST sequences and proteins encoded by them are predicted to have useful biological activities which would make them suitable for treating, preventing or ameliorating medical conditions in humans and animals, although no supporting data is given. Suggested activities include nutritional activity, immune stimulating or suppressing activity, haematopoiesis regulating activity, tissue growth activity, chemotactic/chemokinetic activity, anti-inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, receptor/ligand activity, anti-inflammatory activity, cadherin/tumour invasion suppressor activity, tumour inhibition activity. The EST sequences are also stated to be useful for gene

CC therapy.

XX Sequence 658 BP; 148 A; 176 C; 129 G; 205 T; 0 other;

Alignment Scores: 1.46e-42 Length: 658

Pred. No.: 591.50 Matches: 118

Score: 97.54% Conservative: 1

Best Local Similarity: 96.72% Mismatches: 2

Query Match: 44.84% Indels: 1

DB: 20 Gaps: 1

US-09-987-755-2 (1-249) x AAV87516 (1-658)

Qy 91 ArgArgAlaLysGlyLysAspGlnThrSerSerHisAsnSerSerAspLysAsnArg 110

Db 375 AGGAGAGCCAAAGGGAAGACCGTCATCCCAATCTTCTGATGACAAATCGA 316

Qy 111 ArgAsnSerSerGluGluArgSerArgProAsnSerGlyAspGluLysArgLysLeuSer 130

Db 315 CGTAATTCAGTCAAGAGAGAGAGTAGGCCAACTCAGGTGATGAGAAAGCGCAACTTAGC 256

Qy 131 LeuSerGluGlyLysValLysLysAsnMetGlyGluGlyLysLysArgValSerSerGly 150

Db 255 CTGTCTGAAGGGAAGGTGAAGAACATGGAGAGGAAGAGAGGGTGTCTTCAGGC 196

Qy 151 SerSerGluArgGlySerLysSerProLeuLysArgAlaGlnGluGlnSerProArgLys 170

Db 195 TCTTCAGAGAGAGGCTCCAAATCCCTCTGAAAGAGCCCAAGAGCAAGTCCCGGAAG 136

Qy 171 ArgGlyArgProLysAspGluLysAspLeuThrIleProGluSerSerThrValLys 190

Db 135 CCGGGTCGGCCCCCAAGGATGAGAAGGATCTCAGCATCCCGAGTCTAGTACCGTGAAG 76

Qy 191 GlyMetMetAlaGlyProMetAlaAlaPheLysTrpGlnProThrAla---SerGluPro 209

Db 75 GGGATGATGGCCGACCGATGGCCGCGTAAATGGGAGCCCAACCGCTCTCTAGGCCA 16

Qy 210 ValLys 211

Db 15 TGAAG 10

RESULT 12

AAK70375

ID AAK70375 standard; DNA; 2650 BP.

XX

AC AAK70375;

XX

DT 06-NOV-2001 (first entry)

XX

DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:25187.

XX

KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;

KW cystostatic; gene therapy; vaccine; metastasis; ds.

XX

OS Homo sapiens.

XX

PN WO200157182-A2.

XX

PD 09-AUG-2001.

XX

PF 17-JAN-2001; 2001WO-US01354.

XX

PR 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184684.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 23-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 23-SEP-2000; 2000US-0234997.
PR 23-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.

PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 03-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-483426/52.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides, useful for preventing, diagnosing and/or treating cancers and metastasis -

Disclosure; SEQ ID NO 25187; 307lpp + Sequence Listing; English.

AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I) amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic activity, and can be used in gene therapy and vaccine production. (I) proteins and polynucleotides may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate (I) expression. For example, they may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome

CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK67694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 2650 BP; 754 A; 509 C; 779 G; 608 T; 0 other;

Alignment Scores:
Pred. No.: 3.15e-34 Length: 2650
Score: 504.00 Matches: 119
Percent Similarity: 39.24% Conservative: 5
Best Local Similarity: 37.66% Mismatches: 8
Query Match: 38.21% Indels: 184
DB: 22 Gaps: 1

US-09-987-755-2 (1-249) x AAK70375 (1-2650)

Qy 62 ProTyrHisAlaHisLysGluMetLysLeuLysLeuLysGlyLysArgPheGlnGln 81
Db 2 CCATATCATGCTCATAAAGAGAAATGATAAAATTAACAAGGGTAAACGATTCCAGCA 61
Qy 82 AlaValAspAlaValGluLysPheLeuArgAlaLysGlyLysAsp----- 97
Db 62 GCGGTAGATGCTGTCGAAGAGTTCTCTCAGGAGAGCCAAAGGGAAGACCAAGTGAGAGAC 121
Qy 97 ----- 97
Db 122 ACAATGTTCACTTTCTCAATGCCAGACTGCTGATTTTTCATAGCAACTAACAGATTA 181
Qy 97 ----- 97
Db 182 AATGTCACCCAAACGAAACCTTTGTTCTTCTGAAATGCTACTGGAGAAAGCTACTTCAG 241
Qy 97 ----- 97
Db 242 TTCTAAGGTGTCAGCGTTTGTGTTATCAGTTTGAAACTCCATGCAGCTAAACCTTTCTGGA 301
Qy 97 ----- 97
Db 302 GTAGTGTGTTTAGAAAAATTTGCTTGGACCAGGAGAAATATCTCAGCATTTTGAGT 361
Qy 97 ----- 97
Db 362 CTAGCACAGGCGATTTCTTCTTTCATAGCACACAGATTAGGAGCATAGCCTCTGAATA 421
Qy 97 ----- 97
Db 422 CATATAGCTAGATTTAAATCTTTTCTTAGTTTTGGACATAGGAAAGTACTTT 481
Qy 97 ----- 97
Db 482 TGATATATATTTTTTAAATCTGCTTCTCCCAATGCTCTGGACACAGCAGCAATCTATT 541
Qy 97 ----- 97
Db 542 TCTAGGGTCATTGTGAGGAGACCACACAGAGTGCCTGTAGCAGCACATGGCAGCCTGTAAA 601
Qy 97 ----- 97
Db 602 TGATGGTGGCATGGCGGAGCAGCTGTGGTAGTTATCATTTATTTATTAACCTTT 661
Qy 98 GlnThrSerSerHisAsnSerSerAspLysAsnArgArgAsnSerSerGluGluArg 117
Db 662 CAGAGTCATCCCAAAATTTCTTGATGACAAGATCCGCTAATTCAGTGGAGGAGA 721
Qy 118 SerArgProAsnSerGlyAspGluLysArgLysLeuSerLeuSerGlyLysVallys 137
|||||

Db 722 AGTAGGCCAACTCAGGTGATGAGAAGCGCAAACTTAGCCCTGTCTGAAGGGAAGGTGAAG 781
Qy 138 LysAsnMetGlyGluGlyLysLysArgValSerSerGlySerSerGluArgGlySerLys 157
Db 782 AAGAACATGGGAGAGAGAAAGAGAGGGGTGCTTTCAGGCTTTCAGAGAGAGGCTCCAAA 841
Qy 158 SerProLeuLysArgAlaGlnGlnSerProArgLysArgGlyArgProProLysAsp 177
Db 842 TCCCTCTGTAAGAGAGCCCAAGAGCAAAAGTCCCCGGAAGCGGGTCGCCCCCAAGGAT 901
Qy 178 GluLysAspLeuThrIleProGluSerSerThrValLysGlyMetMet 193
Db 902 GAGAAGGTTTGTGTTTCTCTCTGACGCTTCGCGGATTAGTATGATGTCCTT 949
RESULT 13
AAK70376
ID AAK70376 standard; DNA; 2650 BP.
XX
AC AAK70376;
XX
DT 06-NOV-2001 (first entry)
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:25188.
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccines; metastasis; ds.
XX Homo sapiens.
XX WO200157182-A2.
XX
XX 09-AUG-2001.
XX 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0227009.
PR 01-SEP-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.


```

Db 62 GCGGTAGATGCTGTGAAAGAGTTCTTCAGGAGAGCCAAAGGAAAGACACAGGTGAGAGAC 121
Qy 97 ----- 97
Db 122 ACAATGTTTCACATTTCTACCAATGCCAGAGCTGGTATTTTTCATAGCACTAACAGATTA 181
Qy 97 ----- 97
Db 182 AATGTCACCCAAACGAAACCTTTGTTCTTCTGAAATGCTACTGGAGAAAGCTACTTTCAG 241
Qy 97 ----- 97
Db 242 TTCTAAGGTGTGAGGTTTGTCTTATACAGTTTGAACCTCCATGCAGCTAACCTTTCTGGA 301
Qy 97 ----- 97
Db 302 GTAGGTGTTGCTTTTAGGAAATTTGCTTGGACCCAGGAGAAATATCTCAGCATTTGAGT 361
Qy 97 ----- 97
Db 362 CTAGCACAGGCCATTTCTTCTTCATAGCACACAGATTTAGGAGCATAGCCTCTGAAATA 421
Qy 97 ----- 97
Db 422 CATATACTGCTAGATTTAAATCTTTCTTCTAGTTTGGGACATAGGAAAGTGACTTT 481
Qy 97 ----- 97
Db 482 TGATATATATATTTTAAATCTGCTTCTCCCAATGCTGGACACAGCAGCATCTATT 541
Qy 97 ----- 97
Db 542 TCTAGGTCATTTGTGAGGAGACCACACAGAGTGCCTTAGCAGCACATGGCACCTGTAAA 601
Qy 97 ----- 97
Db 602 TGATGTCGGAGTGGCGGAGCAGCTGTGTGGTAGTTATCATATATTATTAACTTT 661
Qy 98 GlnThrSerSerHisAsnSerSerAspAspLysAsnArgArgAsnSerSerGluGluArg 117
Db 662 CAGAGCTCATCCCAATTTCTTCATGACAGAAATCGACGTAATTCAGTGAGGAGAGA 721
Qy 118 SerArgProAsnSerGlyAspGluLysArgLysLeuSerLeuSerGluGlyValLys 137
Db 722 AGTAGGCCAAACTCAGGTGATGAGAGGCGCAAACTTAGCCTCTCTGAAGGGAAGTGAAG 781
Qy 138 LysAsnMetGlyGluGlyLysArgValSerSerGlySerSerGluArgGlySerLys 157
Db 782 AAGAACATGGGAGAGAGGAAAGAGAGGGTGTCTTCAGGCTCTTTCAGAGAGAGGCTCCAAA 841
Qy 158 SerProLeuLysArgAlaGlnGlnSerProArgLysArgGlyValArgProProLysAsp 177
Db 842 TCCCTCTGAAAGAGCCAGAGCAAGTCCCGGAGCGGGTCTCGGCCCCCAAGGAT 901
Qy 178 GluLysAspLeuThrIleProGluSerSerThrValLysGlyMetMet 193
Db 902 GAGAAGGTTTGTCTCTCTGACGCTTCCGGGATTAGTGATGCTCTT 949

```

RESULT 14

AAV86389/c

ID AAV86389 standard; cDNA; 307 BP.

XX

AC AAV86389;

XX

DT 27-APR-1999 (first entry)

XX

DE EST clone AQ73.

XX

Expressed sequence tag; secreted protein; haematopoiesis regulator;
 tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
 chemotaxis; chemokines; haemostasis; gene therapy; thrombolysis;
 receptor; ligand; anti-inflammatory; tumour inhibitor; ds.

XX

OS Homo sapiens.

XX

PN WO9845435-A2.

XX

PD 15-OCT-1998.

XX

PF 10-APR-1998; 98WO-US06954.

XX

PR 10-APR-1997; 97US-0835913.

XX

PA (GEMY) GENETICS INST INC.

XX

PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;

XX

PI Racie LA, Spaulding V, Treacy M;

XX

DR WPI; 1999-070076/06.

XX

New polynucleotides encoding human secreted proteins - derived from
 e.g. human blood, kidney, foetal lung, placenta, testes, brain,

XX

ovary, pituitary, retina and colon cDNA libraries

XX

Claim 1; Page 218; 633pp; English.

XX

This sequence represents an expressed sequence tag (EST), and is a
 polynucleotide of the invention. The polynucleotides of the invention are
 all secreted EST sequences isolated from a variety of human tissue
 sources. The EST sequences and proteins encoded by them are predicted to
 have useful biological activities which would make them suitable for
 treating, preventing or ameliorating medical conditions in humans and
 animals, although no supporting data is given. Suggested activities
 include nutritional activity, immune stimulating or suppressing activity,
 haematopoiesis regulating activity, tissue growth activity,
 activin/inhibin activity, chemotactic/chemokinetic activity,
 and thrombolytic activity, receptor/ligand activity, anti-inflammatory
 activity, cadherin/tumour invasion suppressor activity, tumour inhibition
 activity. The EST sequences are also stated to be useful for gene
 therapy.

XX

SQ Sequence 307 BP; 67 A; 91 C; 62 G; 85 T; 2 other;

Alignment Scores:

Pred. No.:	3.87e-30	Length:	307
Score:	444.00	Matches:	80
Percent Similarity:	97.59%	Conservative:	1
Best Local Similarity:	96.39%	Mismatches:	2
Query Match:	33.66%	Indels:	0
DB:	20	Gaps:	0

US-09-987-755-2 (1-249) x AAV86389 (1-307)

Qy 1 MetAlaAlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTrpPro 20

Db 271 ATGGCGGCTGTGAGTCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCT 212

Qy 21 ProTrpProGlyLysIleValAsnProLysAspLeuLysLysProArgGlyLysLys 40

Db 211 CTTTGGCCAGAAAGATTGTTTATCCACCAAGGACTTGAAGAAACCTCGCGGAAAGAAA 152

Qy 41 CysPhePheValLysPheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeu 60

Db 151 TGCTTCTTGTGAAATTTTGGAAACAGAGATCATGCTCGGATCAAAGTGGAAACAGCTG 92

Qy 61 LysProTrpHisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGln 80

Db 91 AAGCATATCATGCTCATAAAGAGGATGATATAAATTAACNAGGTAACGATTTCAG 32

Qy 81 GlnAlaVal 83

Db 31 CAAGCACTC 23

RESULT 15

AAC21457

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n.model

Run on: December 8, 2003, 11:35:47 ; Search time 2151 Seconds
(without alignments)
2813.489 Million cell updates/sec

Title: US-09-987-755-2
Perfect score: 1319
Sequence: 1 MAASURLGDLVWKLGRYP.....VCYQATTKKLKICEDLLPLR 249

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09987755/runat_08122003_103436_18319/app_query.fasta_1.391
-DB=EST -Qfmt=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bites -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pcpt -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEA=SIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09987755@cgn 1.1 2810 @runat_08122003_103436_18319 -NCPUP=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_esti:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1291	97.9	1201	13	BX331741	BX331741 BX331741
2	1287	97.6	3741	11	BC032855	BC032855 Homo sapi
3	1283	97.3	1039	12	BM080810	BM080810 AGENCOURT
4	1282	97.2	870	13	BQ720854	BQ720854 AGENCOURT
5	1282	97.2	884	13	BUI58924	BUI58924 AGENCOURT
6	1257	95.3	1003	9	AL536040	AL536040 AL536040
7	1244.5	94.4	1005	12	BM080899	BM080899 AGENCOURT
8	1236.5	93.7	993	13	BUS20128	BUS20128 AGENCOURT
9	1232.5	93.4	890	13	BQ918182	BQ918182 AGENCOURT
10	1228.5	93.1	899	13	BQ892009	BQ892009 AGENCOURT
11	1225.5	92.9	3239	11	AK014456	AK014456 Mus muscu
12	1222	92.6	755	12	BI551885	BI551885 603192793
13	1214.5	92.1	813	12	BI691175	BI691175 603314566
14	1214.5	92.1	929	13	BQ934274	BQ934274 AGENCOURT
15	1211.5	91.8	3250	11	AK040841	AK040841 Mus muscu
16	1207.5	91.5	829	13	BUS93961	BUS93961 AGENCOURT
17	1202.5	91.2	705	13	BU708303	BU708303 UI-M-FCO-
18	1164	88.2	778	12	BI600148	BI600148 603246545
19	1160.5	88.0	773	12	BI103792	BI103792 602887976
20	1159	87.9	816	13	BI459800	BI459800 603200852
21	1155.5	87.6	704	13	BQ768940	BQ768940 UI-M-FCO-
22	1151.5	87.3	801	14	CA512495	CA512495 UI-R-FJO-
23	1144.5	86.8	685	10	BB617747	BB617747 BB617747
24	1144	86.7	816	13	BU373180	BU373180 603589670
25	1144	86.7	1138	12	BM906992	BM906992 AGENCOURT
26	1130.5	85.7	682	14	BY735423	BY735423 BY735423
27	1127	85.4	998	10	BB611224	BB611224 BB611224
28	1111.5	84.3	823	12	BI150279	BI150279 602915496
29	1106.5	83.9	815	12	BI767890	BI767890 603061025
30	1093.5	82.9	732	12	BI218947	BI218947 602938548
31	1077.5	81.7	671	12	BI732831	BI732831 603333939
32	1069.5	81.1	659	12	BM947061	BM947061 UI-M-EHOP
33	1064.5	80.7	660	14	BY753847	BY753847 BY753847
34	1052.5	79.8	646	14	BY745429	BY745429 BY745429
35	1043.5	79.1	786	12	BG976825	BG976825 602845727
36	1035	78.5	775	14	CB601170	CB601170 AGENCOURT
37	1026	77.8	861	12	BI689620	BI689620 603316113
38	1018	77.2	710	9	AJ453917	AJ453917 AJ453917
39	1016.5	77.1	595	14	CB584042	CB584042 AMGNNUC:N
40	1002	76.0	702	9	AJ455240	AJ455240 AJ455240
41	989	75.0	671	9	AJ456583	AJ456583 AJ456583
42	985	74.7	821	12	BG870714	BG870714 602732108
43	977	74.1	908	10	BG294400	BG294400 602391505
44	959	72.7	642	14	BY745902	BY745902 BY745902
45	953	72.3	868	13	BU358607	BU358607 603477645

ALIGNMENTS

RESULT 1
BX331741
LOCUS BX331741 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
DEFINITION CDNA clone CS0DC001YL09 5-PRIME, mRNA sequence.
ACCESSION BX331741
VERSION BX331741.1 GI:30335078
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1201)

AUTHORS
TITLE
JOURNAL
COMMENT

Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 7992.r For

more information about this cluster, see

http://www.genoscope.cns.fr/

cgi-bin/cluster.cgi?seq=CS0DC001CF05QPI&cluster=7992.r. Contact :

Feng Liang Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com/ Invitrogen Corporation 1600

Faraday Avenue Genoscope sequence ID : CS0DC001CF05QPI.

FEATURES

source

1..1201

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CS0DC001YL09"

/tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"

/clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo(dT)

primer. Five prime end enriched, double-strand cDNA was

digested with Not I and cloned into the Not I and EcoR V

sites of the pCMVSPORT 6 vector. Library was normalized."

BASE COUNT 349 a 261 c 332 g 240 t 19 others

ORIGIN

Alignment Scores:

Pred. No.: 4, 18e-96 Length: 1201

Score: 1291.00 Matches: 243

Percent Similarity: 100.00%

Best Local Similarity: 99.59%

Query Match: 97.88%

DB: 13 Gaps: 0

US-09-987-755-2 (1-249) x BX331741 (1-1201)

QY 1 MetAlaValSerLeuArgLeuGlyAspLeuValTrpGlyLeuLeuGlyArgTyrPro 20
DB 75 ATGGCGGCTGTGAGTCTCGGCTCGGACCTGGTGTGGGGAACTCGGCGGATATCCT 134
QY 21 ProTrpProGlyLysLeuValAsnProProLysAspLeuLysLeuProArgGlyLysLys 40
DB 135 CCTTGGCAGGAAGATTGTAATCCACCAAGGACCTTGAAAGAACCTCGCGGAAAGAAA 194
QY 41 CysPhePheValLysPheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeu 60
DB 195 TGCCTTCTTTGCAATTTTGGACAGAGATCATGCTGGATCAAAAGTGGACAGCTG 254
QY 61 LysProTyrHisAlaHisGlyGluMetIleLysIleAsnLysGlyLysArgPheGln 80
DB 255 AAGCATATCATGCTCATAAAGAGAAATGATAAAATTAACAGGGTAAACGATTCCAG 314
QY 81 GlnAlaValAspAlaValGluPheLeuArgAlaLysGlyLysAspGlnThrSer 100
DB 315 CAAGCGGTAGATGCTCTGAAGAGTCTCTCAGGAGAGCCAAAGGAAAGACACAGCTCA 374
QY 101 SerHisAsnSerSerAspLysAsnArgArgAsnSerSerGluGluArgSerArgPro 120
DB 375 TCCACAAATTTCTTGATGACAGATCGAGTAAATTCAGTGAGGAGAGTAGGCCA 434
QY 121 AsnSerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMet 140
DB 435 AACTCAGGTGATGAGAACGCGAACTTAGCCTGTCTGAAGGGAAGGTGAAGAAGACATG 494
QY 141 GlyGluGlyLysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeu 160
DB 495 GGAGAGGAAAGAGAGGGTGTCTTCAGGCTCTTCAGAGAGAGGGCTCCAAATCCCTCTG 554
QY 161 LysArgAlaGlnGlnSerProArgLysArgGlyArgProProLysAspGluLysAsp 180

DB 555 AAAAGAGCCCAAGACGAAAGTCCCGGAAGCGGGTGGCCCCCAAGAGATGAGAAGAT 614
QY 181 LeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPhe 200
DB 615 CTCACCATCCCGGAGTCTAGTACCGTGAAGGGGATGATGGCCGACGATGCGCGGTTT 674
QY 201 LysTrpGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisPhe 220
DB 675 AAATGGCAGCAACCGCAGCGAGCCTGTTAAAGATGACAGATCCTCATTTCCATTTTC 734
QY 221 LeuLeuSerGlnThrGluLysProAlaValCysTyrGlnAlaIleThrLysLysLeuLys 240
DB 735 CTGCTAAGCCCAACAGAGAACCCAGCTGTCTGTACAGGCAATCAGCAAGAGTTGAAA 794
QY 241 IleCysGluAsp 244
DB 795 ATATGTGAAGAG 806

RESULT 2

BC032855

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK

COMMENT

3741 bp mRNA linear HTC 04-MAR-2003

Homo sapiens, Similar to RIKEN cDNA 3930401K13 gene, clone

IMAGE:5265561, mRNA.

BC032855

BC032855.1 GI:23958784

HTC.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 3741)

Strausberg, R.

Direct Submission

Submitted (07-JUN-2002) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC Project URL: http://mgc.nci.nih.gov

Contact: MGC help desk

Email: cgapbs-remail.nih.gov

Tissue Procurement: Miklos Palkovits, M.D., Ph.D.

cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki

Toshiyuki and Piero Carninci (RIKEN)

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Institute for Systems Biology

contact: amadan@systemsbiology.org

Anup Madan, Jessica Fahay, Erin Helton, Mark Kettelman, Anuradha

Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov

Series: IRAC Plate: 47 Row: j Column: 23

This clone was selected for full length sequencing because it

passed the following selection criteria: Hexamer frequency ORF

analysis, Similarity but not identity to protein

This clone has the following problem: retained intron.

Location/Qualifiers

1..3741

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:5265561"

/tissue_type="Testis"

/clone_lib="NIH_MGC_97"

/lab_host="DH10B"

/note="vector: pBluescript"

BASE COUNT 940 a 959 c 993 g 849 t

ORIGIN

Alignment Scores:

Pred. No.: 2,74e-95 Length: 3741

Score: 1287.00 Matches: 242
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 99.18% Mismatches: 0
Query Match: 97.57% Indels: 0
DB: 11 Gaps: 0

US-09-987-755-2 (1-249) x BC032855 (1-3741)

QY 1 MetAlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrPro 20
DB 29 ATGCGCGTGTGAGTCTCGCGCTCGCGACTTGGTGTGGGGAAATCTCGCGCATATCTCT 88
QY 21 ProTrpProGlyLysLeuValAsnProProLysAspLeuLysLysProArgGlyLysLys 40
DB 89 CCTTGGCCAGAAAGATTGTTAATCCACCAAGACTTGAAGAAACCTCGCGGAAAGGAA 148
QY 41 CysPhePheValLysPheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeu 60
DB 149 TGTCTTCTTTGTAATTTTGGACACAGAGATCATGCTCGGATCAAAAGTGGAAACAGCTG 208
QY 61 LysProTyrHisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGln 80
DB 209 AACCCATATCATCTCATTAAGAGGAAATGATAAATTAACAAGGGTAAACGATTCAG 268
QY 81 GlnAlaValAspAlaValGluGluPheLeuArgAlaLysGlyLysAspGlnThrSer 100
DB 269 CAAGCGGTAGATGCTGCGAAGAGTCTCTCAGGAGAGCCAAAGGAAAGACCAAGAGCTCA 328
QY 101 SerHisAsnSerSerAspAspLysAsnArgArgAsnSerSerGluGluArgSerArgPro 120
DB 329 TCCCAACAATTTCTTGATGACAAAGATCGACGTAATTCACAGTGGAGAGAGTAGGCCA 388
QY 121 AsnSerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMet 140
DB 389 AACTCAGTGTATGAGAGCCCAACTTAGCTGTCTGAAGGAGAGTGAGAGAGACATG 448
QY 141 GlyGluGlyLysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeu 160
DB 449 GGAGAGGAAAGAGAGGGTGTCTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTG 508
QY 161 LysArgAlaGlnGluGlnSerProArgLysArgGlyArgProProLysAspGluLysAsp 180
DB 509 AAAAGAGCCCAAGAGCAAGTCCCGGAGCGGGTCCGCCCAAGAGGATGAGAGGAT 568
QY 181 LeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPhe 200
DB 569 CTACCATCCCGAGTCTAGTACCGTGAAGGGATGATGCCGAGCCGATGGCGCGTTT 628
QY 201 LysTrpGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisPhe 220
DB 629 AAATGGCAGCCCAACCGCAAGCGAGCTGTTAAAGATGCAGATCTCATTTCCATTTTC 688
QY 221 LeuLeuSerGlnThrGluLysProAlaValCysTyrGlnAlaIleThrLysLysLeuLys 240
DB 689 CTGCTAAGCCCAACAGAGAGCCAGCTGTCTGTGTACCAGGCAATCACGAAAGAGTTGAAA 748
QY 241 IleCysGluAsp 244
DB 749 ATATGTGAAGAG 760

RESULT 3
BM808810
LOCUS
DEFINITION ACENOCURT 661726 NIH_MGC_124 Homo sapiens cDNA clone IMAGE:5734324
5', mRNA sequence.
ACCESSION BM808810
VERSION BM808810.1 GI:19125633
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1039)

AUTHORS NIH-MGC <http://mgc.mci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Invitrogen
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: Agencourt Bioscience Corporation
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM12739 row: h column: 05
High quality sequence stop: 715.

FEATURES
Location/Qualifiers
1..1039
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5734324"
/tissue_type="hippocampus"
/lab_host="DH10B"
/clone_lib="NIH MGC_124"
/notes="Organ: b7ain; Vector: pCMV-SPORT6; Site_1: EcoRV
(destroyed); Site_2: NotI; RNA source male hippocampus,
age 27. Library is oligo-dT primed and directionally
cloned (EcoRV site is destroyed upon cloning). Average
insert size 1.4 kb, insert size range 0.9-4 kb. Library is
normalized and enriched for full-length clones and was
constructed by C. Gruber (Invitrogen). Research Genetics
tracking code 012."

BASE COUNT 307 a 247 c 288 g 197 t
ORIGIN

Alignment Scores:
Pred. No.: 1,64e-95 Length: 1039
Score: 1283.00 Matches: 241
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 99.18% Mismatches: 0
Query Match: 97.27% Indels: 0
DB: 12 Gaps: 0

US-09-987-755-2 (1-249) x BM808810 (1-1039)

QY 2 AlaAlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrProPro 21
DB 9 GCGCTGTGAGTCTCGCGCTCGCGACTTGGTGTGGGGAAATCTCGCGCATATCTCTCT 68
QY 22 TrpProGlyLysLeuValAsnProProLysAspLeuLysLysProArgGlyLysLysCys 41
DB 69 TGGCCAGGAAAGATTGTTAATCCACCAAGACTTGAAGAAACCTCGCGGAAAGAAATGC 128
QY 42 PhePheValLysPhePheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeuLys 61
DB 129 TTCTTTGTGAATTTTTTGGACACAGAGATCATGCTCGGATCAAAAGTGGAAACAGCTGAAG 189
QY 62 ProTyrHisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGlnGln 81
DB 189 CCATATCATGCTCATTAACAAGGAAATGATAAATTAACAAGGGTAAACGATTCACGAA 248
QY 82 AlaValAspAlaValGluGluPheLeuArgAlaLysGlyLysAspGlnThrSerSer 101
DB 249 GCGGTAGATGCTGCGAAGAGTTCTCTCAGGAGAGCCAAAGGAAAGACCAAGAGCTCATCC 308
QY 102 HisAsnSerSerAspAspLysAsnArgArgAsnSerSerGluGluArgSerArgProAsn 121
DB 309 CACAATTTCTGATGACAAAGATTCACGCTAATTCAGTGGAGAGAGTAGGCCAAAC 368
QY 122 SerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMetGly 141
DB 369 TCAGGTGATGAGAGGAGGCAACATTACCTGTCTGAGGGAGAGGTGAGAGACATGGGA 428
QY 142 GluGlyLysLysArgValSerSerGlySerGluArgGlySerLysSerProLeuLys 161

```

Db 429 GAAGGAAAGAGAGGGTCTCTTACGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTGAAA 488
Qy 162 ArgAlaGlnGluInSerProArgLysArgGlyArgProProLysAspGluLysAspLeu 181
Db 489 AGAGCCCAAGAGCAAGTCCCGGAAGCGGGTCGGCCCCCAAGAGATGAGAGAGATCTC 548
Qy 182 ThrileProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPheLys 201
Db 549 ACCATCCCGGAGTCTAGTACCGTGAGGGGATGATGGCGGACCATGCGCGCTTTAAA 608
Qy 202 TrpGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisPheLeu 221
Db 609 TGGCAGCAACCGCAAGCAGGCTGTAAAGATGACAGATCCTCATTTCCATCATTTCTG 668
Qy 222 LeuSerGlnThrGluLysProAlaValCysTyrGlnAlaIleThrLysLysLeuLysIle 241
Db 669 CTAAGCCCAACAGAGAGCCAGCTGTCTGTACCAAGCAATCACGAAGAAGTTGAATA 728
Qy 242 CysGluAsp 244
Db 729 TGTGAAGAG 737

```

```

RESULT 4
BQ720854
LOCUS
DEFINITION
AGENCOURT 8241262 Lupski sympathetic_trunk Homo sapiens cDNA clone
IMAGE:6187142 5', mRNA sequence.
ACCESSION
BQ720854
VERSION
BQ720854.1 GI:21859751
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgi.nci.nih.gov/.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: Dr. James R. Lupski
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13580 row: k column: 15
High quality sequence stop: 668.
FEATURES
Location/Qualifiers
1..870
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6187142"
/sex="male"
/tissue_type="sympathetic trunk"
/dev_stage="adult, 16 yr"
/lab_host="DH10B"
/clone_lib="Lupski sympathetic trunk"
Note: Site 2: Salt; cDNA made by oligo-dT priming.
Directionally cloned using the following adapters:
5'-TCGACCCACGCGTCCG-3' and
5'-GACTAGTTCTAGATCGGCGCGCGCCCT(15)-3'. Size selected >
1 kb for average insert length 1.9 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine); available through Life
Technologies."
BASE COUNT
270 a 193 c 245 g 161 t 1 others
ORIGIN

```

```

Alignment Scores:
Pred. No.: 1.67e-95 Length: 870
Score: 1282.00 Matches: 241
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.59% Mismatches: 0
Query Match: 97.19% Indels: 0
DB: 13 Gaps: 0
US-09-987-755-2 (1-249) x BQ720854 (1-870)

```

```

Qy 3 AlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrProProTrp 22
Db 3 GCTGTAGTCTCGGCTCGGCGACTTGTGTGGGGAAACTCGGCCGATATCTCTCTTGG 62
Qy 23 ProGlyLysIleValAsnProProLysAspLeuLysLysProArgGlyLysLysCysPhe 42
Db 63 CCAGGAAGATTGTGTAATCCCAAGGACTTGAAGAAACCTCGCGGAAGAAATGCTTC 122
Qy 43 PheValLysPhePheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeuLysPro 62
Db 123 TTTGTGAATTTTGGAAACAGAGATCATGCTGGATCAAAGTGGAAACAGCTGAAGCA 182
Qy 63 TyrHisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysA:GpGlnGlnAla 82
Db 183 TATCATGCTCATAAAGAGGAAATGATAAAATTAACAAGGGTAAACGATTCCAGCAAGCG 242
Qy 83 ValAspAlaValGluGluPheLeuArgArgAlaLysGlyLysAspGlnThrSerSerHis 102
Db 243 GTAGATGCTGTCGAAGAGTTCTCTCAGAGAGCCAAAGGGAAAGACCAAGCTATCCAC 302
Qy 103 AsnSerSerAspAspLysAsnArgArgAsnSerSerGluGluArgSerArgProAsnSer 122
Db 303 AATTCTTCTGATGACAGAAATCGACGTAAATTCAGTGGAGAGAGAGTAGGCCAACTCA 362
Qy 123 GlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMetGlyGlu 142
Db 363 GGTGATGAGAAGCGCAAACTTAGCTCTCTGAAGGGAAGGTGAAGAAGAACATGGAGAA 422
Qy 143 GlyLysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeuLysArg 162
Db 423 GGAAAGAAGAGGGGTCTTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCCCTCTGAAAAGA 482
Qy 163 AlaGlnGluInSerProArgLysArgGlyArgProProLysAspGluLysAspLeuThr 182
Db 483 GCCCAAGAGCAAACTCCCGGAAGCGGGTCCGCCCCCAAGAGATGAGAGAGATCTCACC 542
Qy 183 IleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPheLysTrp 202
Db 543 ATCCCGGAGTCTAGTACCGTGAAGGGATGATGCGCGGACCGATGCGCGCTTTAAATGG 602
Qy 203 GlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisPheLeuLeu 222
Db 603 CAGCCAAACCGCAAGCAGCGCTGTAAAGATGACAGATCTCTATTTCCATCATTTCTTGCTA 662
Qy 223 SerGlnThrGluLysProAlaValCysTyrGlnAlaIleThrLysLysLeuLysIleCys 242
Db 663 AGCCAAACAGAGAGCCAGCTGTCTGTACCGCAATCACGAAGAAGTTGAATAATATGT 722
Qy 243 GluAsp 244
Db 723 GAAGAG 728

```

```

RESULT 5
BUI58924
LOCUS
DEFINITION
AGENCOURT 7844927 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6044084
5', mRNA sequence.
ACCESSION
BUI58924
VERSION
BUI58924.1 GI:22672834
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 884)
 NIH-MGC http://mgi.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L14M13286 row: b column: 21
 High quality sequence stop: 776.

FEATURES

source
 1. .884
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6044084"
 /tissue_type="embryonal carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH MGC 92"
 /note="Organ: testis; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 2.5 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH MGC Library."
 BASE COUNT 274 a 195 c 244 g 169 t 2 others
 ORIGIN

Alignment Scores:
 Pred. No.: 1.69e-95 Length: 884
 Score: 1282.00 Matches: 241
 Percent Similarity: 100.00% Conservatives: 1
 Best Local Similarity: 99.59% Mismatches: 0
 Query Match: 97.19% Indels: 0
 DB: 13 Gaps: 0

US-09-987-755-2 (1-249) x BU158924 (1-884)

Qy 3 AlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTrpProTyr 22
 Db 2 GCTGTGAGTCGCGGCTCGCGGCACTTGTGTGGGGGAACTCGCGGCAATCTCTCTGG 61
 Qy 23 ProGlyLysLeuValAsnProProLysAspLeuLysLysProArgGlyLysLysCysPhe 42
 Db 62 CCAGGAAAGATTGTTAATCCACCAAGGACCTTGAAGAAACCTCGCGGAAAGAAATGCTTC 121
 Qy 43 PheValLysPheGlyThrGluAspHisAlaTrpLysValGluGlnLeuLysPro 62
 Db 122 TTTGTGAATTTTGTGAACAGAGATCATCTCGGATCAAAGTGAAGTGAACGCA 181
 Qy 63 TyrHisAlaHisLysGluGluMetIleLysLysLysLysLysLysLysLysLysLysLys 82
 Db 182 TATCATGCTCATAAGAGGAATGATTAATAATTAACAGGGTAACGATTCACCAACGCG 241
 Qy 83 ValAspAlaValGluGluPheLeuArgAlaLysGlyLysLysLysLysLysLysLysLys 102
 Db 242 GTAGATGCTGCGAAGAGTCTCTCAGGAGACCAAGGAAAGGAAAGGAGTCAATCCAC 301
 Qy 103 AsnSerSerAspAspLysAsnArgArgAsnSerSerGluGluArgSerArgProAsnSer 122
 Db 302 AATTCTCTGATCAAGATCGATTAATTCAGTCAGGAGAGAGTAGGCCAACTCA 361
 Qy 123 GlyAspGluLysArgLysLeuSerLysSerGluGlyLysValLysLysLysLysLysLys 142
 Db 362 GGTATGAGAGCGCAACTTAGCTGTCTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 421
 Qy 143 GlyLysLysArgValSerSerGlySerGluArgGlySerLysSerProLeuLysArg 162

Db 422 GGAAGAAGAGGGGTGCTTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTGAAAGA 481
 Qy 163 AlaGlnGlnSerProArgLysArgGlyArgProProLysAspGlyLysAspLeuThr 182
 Db 482 GCCCAGAGCAGTCCCGGAGCGGGTGGCCCCCAAGAGGATGAGAGGATCTCACC 541
 Qy 183 IleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPheLysTyr 202
 Db 542 ATCCCGGAGTCTAGTACCGTGAAGCGGATGATGCGCGGACCGATGCGCGGTTAAATGG 601
 Qy 203 GlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisPheLeuLeu 222
 Db 602 CAGCCCAACCGCAGCGGAGCTGTAAAGATGCGAGATCTTCATTCATTCATTCCTGCTA 661
 Qy 223 SerGlnThrGluLysProAlaValCysTyrGlnAlaIleThrLysLysLysLysLys 242
 Db 662 AGCCAAACAGAGAGCGGCTGTCTGTACAGGCAATCAGAGAGAGTTGAAATATGT 721
 Qy 243 GluAsp 244
 Db 722 GAAGAG 727

RESULT 6

AL536040
 LOCUS
 DEFINITION AL536040 Homo sapiens FETAL BRAIN Homo sapiens linear EST 31-MAY-2003
 CS0DF022Y005 5-PRIME, mRNA sequence.
 ACCESSION AL536040
 VERSION AL536040.2 GI:31260910
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1003)
 Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 Full-length cDNA libraries and normalization
 JOURNAL Unpublished
 COMMENT On Feb 13, 2001 this sequence version replaced gi:12799533.
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of
 Invitrogen. This sequence belongs to sequence cluster 7982.r For
 more information about this cluster, see
 http://www.genoscope.cns.fr/
 cgi-bin/cluster.cgi?seq=CS0DF022AH03QPI&cluster=7982.r. Contact :
 Feng Liang Email : fliang@lifetech.com URL :
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
 Faraday Avenue Genoscope sequence ID : CS0DF022AH03QP1.

FEATURES

source
 1. .1003
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DF022Y005"
 /tissue_type="FETAL BRAIN"
 /dev_stage="fetal"
 /clone_lib="Homo sapiens FETAL BRAIN"
 /notes="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."
 BASE COUNT 308 a 221 c 273 g 196 t 5 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 2.16e-93 Length: 1003
 Score: 1257.00 Matches: 240
 Percent Similarity: 98.37% Conservatives: 1


```

Qy 161 LysArgAlaGlnGluInSerProArgLysArgGlyArgProProLysAspGluLysAsp 180
Db 500 AAAAGAGCCCAAGAGCAAGTCCCGGAGCGGGTCGGCCCAAGAGGATGAGAAGAT 559

Qy 181 LeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPhe 200
Db 560 CTCACCATCCCGGAGTCTAGTACCGTGAAGGGGATGATGGCCGACCGATGGCCGCGTTT 619

Qy 201 LysTrpGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisPhe 220
Db 620 AAATGGAGCAACCGCAAGGAGCGCTGTAAAGATGCAGATCTCATTTCCATCATTTTC 679

Qy 221 LeuLeuSerGlnThrGluLysProAlaValCysTyrGlnAlaIleThrLysLysLeu--- 239
Db 680 CTGCTAAGCCCAACAGAGAGCCAGCTGTCTGTACCAAGGCGCAATCACCAGGAAGTGA 739

Qy 240 LysIleCysGlu 243
Db 740 AAATTATGTGAA 751

RESULT 8
BU520128 993 bp mRNA linear EST 12-SEP-2002
LOCUS AGENCOURT 10157423 NIH_MGC_134 Mus musculus cDNA clone
DEFINITION IMAGE:6518047 5', mRNA sequence.
ACCESSION BU520128
VERSION BU520128.1 GI:22827654
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
JOURNAL NIH-MGC http://mgc.nci.nih.gov/.
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. David Rowe
cDNA Library Preparation: Invitrogen Corp
DNA Sequencing by: The I.M.A.G.E. Consortium (LMNL)
Distribution by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMNL at:
http://image.llnl.gov
Plate: LLM44099 row: c column: 08
High quality sequence stop: 675.
Location/Qualifiers
1. 993
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6518047"
/tissue_type="undifferentiated limb"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_134"
/notes="Vector: pCMV-SPORT6.1.cdb; Site 1: EcoRV; Site 2:
NotI; Cloned unidirectionally. Primer: Oligo dt. Average
insert size 1.7 kb. Constructed by ResGen, Invitrogen
Corp. Note: this is a NIH_MGC Library."

BASE COUNT 293 a 222 c 275 g 194 t 9 others
ORIGIN

Alignment Scores:
Pred. No.: 1.03e-91 Length: 993
Score: 1236.50 Matches: 233
Percent Similarity: 98.35% Conservative: 6
Best Local Similarity: 95.88% Mismatches: 3
Query Match: 93.75% Indels: 1
DB: 13 Gaps: 1

US-09-987-755-2 (1-249) x BU520128 (1-993)

```

```

Qy 2 AlaAlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrProPro 21
Db 1 CGGCTGTGAGTCTCGGCTCGGCGACTTGGTGTGGGGAAACTGGGCGGTATCTCTCCC 60

Qy 22 TrpProGlyLysIleValAsnProProLysAspLeuLysLysProArgGlyLysLysCys 41
Db 61 TGGCCAGGAAAGATTTGTTAATCCACCCAGGACTTAAAGAAACCCACCTGGAAAGAAATGC 120

Qy 42 PhePheValLysPhePheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeuLys 61
Db 121 TTCTTTGTGAAGTTTTTGGAAACAGAGATCATGCTTGGATCAAAAGTGGACACGCTAAAG 180

Qy 62 ProTyrHisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGlnGln 81
Db 181 CCTTACCATGCTCACAGGAGGAGATGATAAGATTAAACAGGGTAAACGGTTCACGCAA 240

Qy 82 AlaValAspAlaValGluGluPheLeuArgArgAlaLysGlyLysAspGlnThrSerSer 101
Db 241 GCTGTGGATGCTGTGAAGAGTTCTCTCAGGAGAGCCAAAGGGAAGACACGACATCATCC 300

Qy 102 HisAsnSerSerAspAspLysAsnArgArgAsnSerSerGluGluArgSerArgProAsn 121
Db 301 CACATTCTCTGATGACAAAGATCGCGGTAATTCCAGTGAGGAGAGAGTAGGCCCAAC 360

Qy 122 SerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMetGly 141
Db 361 TCAGGTGATGAGAAACCGAAGCTTAGCTCTCTGAAGGGAAGGTGAAGAGAGACATCGGA 420

Qy 142 GluGlyLysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeuLys 161
Db 421 GAAGGAAAGAAAGAGGGTGCATCTCAGGCTCTGCAGACACAGAGGCTCCAAATGC---CTTAAA 477

Qy 162 ArgAlaGlnGlnGlnSerProArgLysArgGlyArgProProLysAspGluLysAspLeu 181
Db 478 AGAGCCCAAGAGCAAGTCCCCGGAAGCGGGTGGCCCCCAAGGATGAGAGAGGAGCTC 537

Qy 182 ThrIleProGluSerSerThrValLysGlyMetAlaGlyProMetAlaAlaPheLys 201
Db 538 ACCATCCCTGAGTCTAGCACTGTAAAGGGATGATGGTGGACCGATGGCTGCATTANA 597

Qy 202 TrpGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisPheLeu 221
Db 598 TGGCAGCAACACAGCCGACCGAGCAGTCAAGATGCAGATCCTCATTTTCATCTTTCTG 657

Qy 222 LeuSerGlnThrGluLysProAlaValCysTyrGlnAlaIleThrLysLysLeuLysIle 241
Db 658 TTGAGCCAAACAGAGAGCCAGCTGTCTGTACAGGCAATCACAAAGAAAGTTGAAATA 717

Qy 242 CysGluAsp 244
Db 718 TGTGAAGAA 726

RESULT 9
BU520128 890 bp mRNA linear EST 20-AUG-2002
LOCUS AGENCOURT 8765993 NIH_MGC_130 Mus musculus cDNA clone IMAGE:6330861
DEFINITION 5', mRNA sequence.
ACCESSION BU520128
VERSION BU520128.1 GI:22332892
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
JOURNAL NIH-MGC http://mgc.nci.nih.gov/.
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman,
Ph.D.

```

cDNA Library Preparation: ResGen, Invitrogen Corp
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM13785 row: o column: 22
 High quality sequence stop: 648.

FEATURES

1. 890
 Location/Qualifiers
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:6330861"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH MGC_130"
 /note="Organ: Oocytes; Vector: pCMV-SPORT6.1.cdb;
 Site 1: EcorV; Site 2: NotI; Cloned unidirectionally.
 Primer: Oligo dt. Average insert size 1.95 kb.
 Constructed by ResGen, Invitrogen Corp. Note: this is a
 NIH MGC Library."

BASE COUNT 268 a 201 c 245 g 175 t 1 others
 ORIGIN

Alignment Scores:
 Pred. No.: 1,978-91 Length: 890
 Score: 1232.50 Matches: 232
 Percent Similarity: 98.35% Conservative: 6
 Best Local Similarity: 95.87% Mismatches: 3
 Query Match: 93.44% Indels: 1
 DB: 13 Gaps: 1

US-09-987-755-2 (1-249) x BQ918182 (1-890)

Qy 3 AlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyProProTrp 22
 Db 3 GCTGTGAGTCTCGCGCTCGGACCTGTGTGGGGGAAACTGGCGCGGTATCTCTCCCTGG 62
 Qy 23 ProGlyLysIleValAsnProProLysAspLeuLysLysProArgGlyLysLysCysPhe 42
 Db 63 CCAGGAAGATGTTTAATCCCAAGGACCTTAAGAAACCAACGCTGGGAAGAAATGCTTC 122
 Qy 43 PheValLysPhePheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeuLysPro 62
 Db 123 TTTGTGAAGTCTTTTGGACAGAGATCATGCTGTGATCAAGTGAAGTGAACAGCTTAAGGCT 182
 Qy 63 TyrHisAlaHisLysGluMetIleLysIleAsnLysGlyLysArgPheGlnAla 82
 Db 183 TACCATGCTCACAGGAGGATGATAAAGATTAAACAAGGGTAAACGGTTCACGCAAGCT 242
 Qy 83 ValAspAlaValGluGluPheLeuArgAlaLysGlyLysAspGlnThrSerSerHis 102
 Db 243 GTGGATGCTGTGAAGAGTCTCTCAGGAGAGCCAAAGGGAAGACCAACATCATCCAC 302
 Qy 103 AsnSerSerAspLysAsnArgAsnSerSerGluGluArgSerArgProAsnSer 122
 Db 303 ACTTCTGCTGATGACAAATCGCGGTATTCAGTAGGAGAGAGTAGGCCAACTCA 362
 Qy 123 GlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMetGlyGlu 142
 Db 363 GGTGATGAGAAACGCAAGCTTAGCTGTCTGAAGGGAAGGTGAAGAAACATGGGAGA 422
 Qy 143 GlyLysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeuLysArg 162
 Db 423 GGAAGAAGAGGGTGACTTTCAGGCTCTGCACAGAGAGGCTCCAAATGC---CTTAAAGA 479
 Qy 163 AlaGlnGluInSerProArgLysArgGlyArgProProLysAspGluLysAspLeuThr 182
 Db 480 GCCCAAGAGCAAGTCCCGGAGCGGGTGGCCGCCCAAGAGGTAGAGAGGACTTACC 539
 Qy 183 IleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPheLysTrp 202
 Db 540 ATCCCTGAGTCTAGCAGCTGTAAGGGGATGATGGCTGACCGGATGGCTGCTATTAATGG 599

Qy 203 GlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisPheLeuLeu 222
 Db 600 CAGCCACAGACGACGAGCCAGTCAAGATGCAGATCTCATTTTCATCTTCTGTG 659
 Qy 223 SerGlnThrGluLysProAlaValCysTyroGlnAlaIleThrLysLysLeuLysIleCys 242
 Db 660 AGCCAAACAGAGAGCCAGCTGTGTGTACCAAGCAATCAAAAGAGTTTGAATATGT 719
 Qy 243 GluAsp 244
 Db 720 GAAGAG 725

RESULT 10

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

Location/Qualifiers

1. 899

/organism="Mus musculus"

/mol_type="mRNA"

/db_xref="taxon:10090"

/clone="IMAGE:6391577"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH MGC_129"

/note="Organ: Olfactory epithelium; Vector: pCMV-SPORT6.1.cdb; Site 1: EcorV; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this is a NIH MGC Library."

BASE COUNT 277 a 190 c 257 g 174 t 1 others

ORIGIN

Alignment Scores:

Pred. No.: 4,23e-91 Length: 899

Score: 1228.50 Matches: 231

Percent Similarity: 97.95% Conservative: 8

Best Local Similarity: 94.67% Mismatches: 4

Query Match: 93.14% Indels: 1

DB: 13 Gaps: 1

US-09-987-755-2 (1-249) x BQ992009 (1-899)

Qy 1 MetAlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyPro 20

Db 5 ATGCGCGCTGTGAGTCTGCGCTCGGACCTTGTGTGGGGAACTGGGCGCGGTATCCT 64

Qy 21 ProTrpProGlyLysIleValAsnProProLysAspLeuLysLysProArgGlyLysLys 40

Db 65 CCCTGGCCAGAAAGATTTGTTAATCCCAAGGACTTAAAGAAACCAACCGTGGAAAGAAA 124

Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT 236 a 156 c 218 g 145 t
ORIGIN

Alignment Scores:

Pred. No.: 1,22e-90 Length: 755
Score: 1222.00 Matches: 240
Percent Similarity: 97.96% Conservative: 0
Best Local Similarity: 97.96% Mismatches: 3
Query Match: 92.65% Indels: 3
DB: 12 Gaps: 0

US-09-987-755-2 (1-249) x BI551885 (1-755)

```
Qy 1 MetAlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrPro 20
Db 25 ATGGCGGCTGTGAGTCTGCGGCTCGCGACTTGTGTGGGGGAAACTCGGCCGATATCCT 84
Qy 21 ProTyrProGlyLysIleValAsnProProLysAspLeuLysLysProArgGlyLysLys 40
Db 85 CCTTGGCCAGGAAGATTGTATCCACCAAGGACTTGAAGAACTTCGCGGAAGAA 144
Qy 41 CysPhePheValLysPheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeu 60
Db 145 TGCTTCTTGTGAAATT-TTTGGAAACAGAGATCATGCTGGTCAAAAGTGGAAACAGCTG 203
Qy 61 LysProTyrHisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGln 80
Db 204 AAGCCATATCATGCTCATGAAGAGGAATGATAAAATTAAACAGGGTAAACCATTCAG 263
Qy 81 GlnAlaValAspAlaValGluPheLeuArgArgAlaLysGlyLysAspGlnThrSer 100
Db 264 CAAGCGGTAGTGTCTGTGAGAGAGTCTCTCAGGAGAGCCAAAGGAAAGACGACGTCA 323
Qy 101 SerHisSerSerAspAspLysAsnArgArgAsnSerSerGluGluArgSerArgPro 120
Db 324 TCCCACAATCTTCTGATGACAGAAATCGAGTAAATCCAGTGAGGAGAGAGTAGGCCA 383
Qy 121 AsnSerGlyAspGluLysArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMet 140
Db 384 AACTCAGGTGATGAGAGCGCAACTTAGCTGTCTGAAGGGAAGGTGAAGAAAGCAATG 443
Qy 141 GlyGluGlyLysLysArgValSerSerGlySerSerGluArgGlySerLysSerProLeu 160
Db 444 GGAGAGGAAGAAAGAGGGTGTCTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTG 503
Qy 161 LysArgAlaGlnGluGlnSerProArgLysArgGlyArgProProLysAspGluLysAsp 180
Db 504 AAAAGAGCCCAAGAGCAAGTCCCGGAAGCGGGTCCGGCCCAAGAGGATGAGAGGAT 563
Qy 181 LeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPhe 200
Db 564 CTCACCATCCCGGAGTCTAGTACCGTGAAGGGGATGATGCGCCGACCGATGGCGCGTTT 623
Qy 201 LysTrpGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisHis-Ph 220
Db 624 AAATGGGAGCAAAACCGAAGGAGCGCTGTAAAGATGCGAGATCTCATTTCCATCTTT 683
Qy 220 eLeuLeuSerGln-ThrGluLysProAlaValCysTyrGlnAlaIleThrLysLysLeuL 240
Db 684 CCTGCTAAGCCAAACAGAGAGGACGCTGCTGTACCGAGGCAATCAAGAGAGTTGA 743
Qy 240 YsIleCysGlu 243
Db 744 AAATATGTGAA 754
```

RESULT 13

BI691175 813 bp mRNA linear EST 18-SEP-2001
LOCUS 603314566F1 NCI_CGAP_Mam6 Mus musculus cdna clone IMAGE:5354441 5',
DEFINITION mRNA sequence.

ACCESSION

BI691175 GI:15653804

VERSION

EST.

KEYWORDS

Mus musculus (house mouse)

SOURCE

Mus musculus

ORGANISM

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NIH-MGC <http://mgi.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished

Contact: Robert Strausberg, Ph.D.

Email: cgabbs-remail.nih.gov

Tissue Procurement: Jeffrey Green M.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM11901 row: c column: 18

High quality sequence stop: 806.

Location/Qualifiers

1..813

/organism="Mus musculus"

/mol_type="mRNA"

/strain="FVB/N"

/db_xref="taxon:10090"

/clone="IMAGE:5354441"

/sex="female, virgin"

/tissue_type="infiltrating ductal carcinoma"

/dev_stage="5 months"

/lab_host="DH10B"

/clone_lib="NCI CGAP Mam6"

/notes="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;

Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.

Library constructed by Life Technologies. Investigator

providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 256 a 179 c 227 g 151 t

ORIGIN

Alignment Scores:

Pred. No.: 5.4e-90 Length: 813
Score: 1214.50 Matches: 234
Percent Similarity: 97.55% Conservative: 5
Best Local Similarity: 95.51% Mismatches: 3
Query Match: 92.08% Indels: 3
DB: 12 Gaps: 1

US-09-987-755-2 (1-249) x BI691175 (1-813)

```
Qy 1 MetAlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyrPro 20
Db 2 ATGGCGGCTGTGAGTCTGCGGCTCGCGACTTGTGTGGGGGAAACTCGGCCGATATCCT 61
Qy 21 ProTyrProGlyLysIleValAsnProProLysAspLeuLysLysProArgGlyLys-Ly 40
Db 62 CCTTGGCCAGGAAGATTGTATCCACCAAGGACTTAAAGAAACCCACGTGGAAACGAA 121
Qy 40 sCysPhePheValLysPhePheGlyThrGluAspHisAlaTrpIleLysValGluGlnLe 60
Db 122 ATGCTCTTGTGAAGTTTTTGGAAACAGAGATCATGCTCGTGGATCAAGTGAACAGCT 181
Qy 60 uLysProTyrHisAlaHisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGln 80
Db 182 AAAGCCTTACCATGCTCACAAGGAGGAGATGATAAAGATTAAACAGGGTAAACGGTTCCA 241
Qy 80 nGlnAlaValAspAlaValGluGluPheLeuArgArgAlaLysGlyLysAspGlnThrSe 100
Db 242 GCAGCTGTGAGTCTGTGTAAGAGTTCTTCAGAGAGGCCAAAGGAAAGACACGACATC 301
Qy 100 rSerHisAsnSerSerAspAspLysAsnArgArgAsnSerSerGluGluArgSerArgPr 120
```

```

Db      302 ATCCACACATTCCTGCTGATGACAAGATCGCGTAAATTCAGTGAGGAGAGAAGTAGGCC 361
Qy      120 oasnSerGlyAspGluLysAspGlyLysSerLeuSerGlyLysValLysValLysAsnMe 140
Db      362 AAACCTCAGGTGATGAAACGCAAGCTTAGCCTGCTGAAGGAAGGTGAAGAAGAAAT 421
Qy      140 tGlyGluGlyLysLysArgValSerSerGlySerSerGlySerGluArgGlySerLysSerProLe 160
Db      422 GGGAGAAGAAAGAGAGGCTGACTTCAGGCTCTGCACAGACAGAGCTCCAAATGC--CT 478
Qy      160 uLysArgAlaGlnGlnGlnSerProArgLysArgGlyArgProProLysAspGluLysAs 180
Db      479 TAAAGAGAGCCCAAGAGCAAGTCCCGGAAGCGGGTCGGCCCCCAAGAGGATGAGAAGGA 538
Qy      180 pleuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPh 200
Db      539 CTCACCATCCCTGAGCTAGACACTGTAAAGGGGATGATGGCTGCAGCATGGCTGCATT 598
Qy      200 elystrpGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisPhe 220
Db      599 TAAATGGCAGCCCAACAGCGACCGAGCCAGTCAAAAGATCAGATCCTCATTTTCATATT 658
Qy      220 eleuLeuSer-GlnThrGluLysProAlaValCysTy-GlnAlaIleThrLysLysLeuL 240
Db      659 TCTGTTGAGCCCAACAGAGAGGAGGCTGTCTGTATCCAGGCAATCAAAAGAGTTGA 718
Qy      240 yelleCysGlu 243
Db      719 AAATATGTGAA 729

```

```

RESULT 14
BO934274
LOCUS      BO934274
DEFINITION AGENCOURT_8752717 NIH_MGC_130 Mus musculus cDNA clone IMAGE:6335698
5', mRNA sequence.
ACCESSION BO934274
VERSION   BO934274.1 GI:22349657
KEYWORDS  EST.
SOURCE    Mus musculus (house mouse)

```

```

ORGANISM  Mus musculus
REFERENCE 1
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-r@mail.nih.gov
          Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman, Ph.D.

```

```

cDNA Library Preparation: ResGen, Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM13798 row: i column: 11
High quality sequence stop: 701.
Location/Qualifiers
1..929
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6335698"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_130"
/notes="Organ: Otcysts; Vector: pCMV-SPORT6.1.ccdB; Site 1: EcoRV; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.95 kb. Constructed by ResGen, Invitrogen Corp. Note: this is a NIH_MGC Library."

```

```

FEATURES
source

```

```

BASE COUNT 285 a 198 c 258 g 182 t 6 others
ORIGIN

```

Alignment Scores:

```

Pred. No.: 6.17e-90 Length: 929
Score: 1214.50 Matches: 228
Percent Similarity: 97.91% Conservatives: 6
Best Local Similarity: 95.40% Mismatches: 4
Query Match: 92.08% Indels: 1
DB: 13 Gaps: 1

```

```

US-09-987-755-2 (1-249) x BO934274 (1-929)

```

```

Qy      6 LeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyProProTrpProGlyLys 25
Db      2 CTGCGGCTCGCGACTGGTGTGGGGAAACTGGGGCGGTATCTCTCCCTGCCAGGAAG 61
Qy      26 IleValAsnProProLysAspLeuLysLysProArgLysLysLysCysPheValLys 45
Db      62 ATTGTATATCCACCAAGGACTTAAAGAAACCACTGGGAAAGAAATGCTTTTGTGAAG 121
Qy      46 PhePheGlyThrGluAspHisAlaTrpIleLysValGluGlnLeuLysProTyHisAla 65
Db      122 TTTTGTGAACAGAGATCATGCTTGGATCAAGTGGNACAGCTTAAGCCTTACCATGCT 181
Qy      66 HisLysGluGluMetIleLysIleAsnLysGlyLysArgPheGlnGlnAlaValAspAla 85
Db      182 CACAAGGAGGAGATGATAAGATTAAACAGGGTAAACGGTTCCAGCAAGCTGTGGATGCT 241
Qy      86 ValGluGluPheLeuArgArgAlaLysGlyLysAspGlnThrSerSerHisAsnSerSer 105
Db      242 GTTGAAGAGTTCTCTCAGGAGAGCCAAAGGGGAAAGACAGACATCATCCACACTTCTGCT 301
Qy      106 AspAspLysAsnArgArgAsnSerSerGluGluArgSerArgProAsnSerGlyAspGlu 125
Db      302 GATCACAAGATCGCGTATTCAGTGAGGAGAGAGTAGGCCCAACTCAGGTGATGAG 361
Qy      126 LysArgLysLeuSerLeuSerGluGlyLysValLysLysAsnMetGlyGluGlyLysLys 145
Db      362 AAACGCAAGCTTAGCTGTCTGAAGGGAAGGTGAAGAAACATCGGGAAGGAAGGAAG 421
Qy      146 ArgValSerSerGlySerSerGluArgGlySerLysSerProLeuLysArgAlaGlnGlu 165
Db      422 AGGTGTAGCTTCAGGCTCTGCAGACAGAGAGCTCCAAATGC---CTTAAAGAGGCCAAGAG 478
Qy      166 GlnSerProArgLysArgGlyArgProProLysAspGlyLysAspLeuThrIleProGlu 185
Db      479 CAAAGTCCCCGAAGCGGGGTGGCCCCCAAGAGATGAGAAGGACCTCACCATCCCTGAG 538
Qy      186 SerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPheLysTrpGlnProThr 205
Db      539 TCTAGCACTGTAAAGGGGATGATGGCTGGACCGATGCTGCATTTANATGGCAGCAACA 598
Qy      206 AlaSerGluProValLysAspAlaAspProHisPheHisPheLeuLeuSerGlnThr 225
Db      599 GCGAGCCGAGCGAGTCAAAAGATGCGAGATCCTCATTTTTCATCATTTTCTGTGAGCAACA 658
Qy      226 GlutLysProAlaValCysTyGlnAlaIleThrLysLysLeuLysLysCysGluAsp 244
Db      659 GAGAGCGAGCTGTCTGTATCCAGGCAATCAAAAGAGTTGAANATATGTGAAGAG 715

```

```

RESULT 15
AK040841
LOCUS

```

```

DEFINITION

```

```

ACCESSION

```

```

VERSION

```

```

KEYWORDS

```

```

SOURCE

```

```

ORGANISM

```

```

REFERENCE

```

```

AK040841 3250 bp mRNA linear HTC 05-DEC-2002
Mus musculus adult male aorta and vein cDNA, RIKEN full-length
enriched library, clone: A530029N17 product: CYTOKINE-LIKE NUCLEAR
FACTOR N-PAC homolog [Homo sapiens], full insert sequence.

```

```

AK040841
AK040841.1 GI:26088102
HTC; CAP trapper.
Mus musculus (house mouse)

```

```

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

AUTHORS

TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 9279253
PUBMED 10349636

REFERENCE

AUTHORS

Carninci, P., Shibata, Y., Matsuo, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitagawa, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsuo, N., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuo, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multipipillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nika, I., Pesole, G., Quackenbush, J., Schriml, L., Stauber, E., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M., Brownstein, M., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D., Kamiya, M., Lee, N., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombauts, P., Nordone, P., Ring, P., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K., Suzuki, H., Toyooka, K., Wang, K., Weitz, C., Whittaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohtsuki, S., and Hayashizaki, Y.

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hasegawa, T., Hara, A., Hashizume, M., Hayashida, K., Hayashizaki, Y., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koyama, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.

Direct Submission

Submitted (16-JUL-2001)

Yoshihide Hayashizaki, The Institute of

Physical and Chemical Research (RIKEN), Laboratory for Genome

Exploration Research Group, RIKEN Genomic Sciences Center (GSC),

Miyazaki, Japan

11042159

11076861

11217851

COMMENT

RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/.

FEATURES

source

1. 3250
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:AS30029N17"
/db_xref="taxon:10090"
/clone="AS30029N17"
/sex="male"
/tissue type="aorta and vein"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"

misc_feature

19. 1658
/notes="CYTOKINE-LIKE NUCLEAR FACTOR N-PAC homolog [Homo sapiens] (SPT93BXX2, evidence: FASTY, 97.8%ID, 100%length, match=1638)
putative"

BASE COUNT 831 a 783 c 846 g 790 t

ORIGIN

Alignment Scores:

Pred. No.: 3,75e-89 Length: 3250
Score: 1211.50 Matches: 233
Percent Similarity: 97.95% Conservative: 6
Best Local Similarity: 95.49% Mismatches: 4
Query Match: 91.85% Indels: 2
DB: 11 Gaps: 1

US-09-987-755-2 (1-249) x AK040841 (1-3250)

Qy 1 MetAlaAlaValSerLeuArgLeuGlyAspLeuValTrpGlyLysLeuGlyArgTyPro 20

Db 19 ATGGCGCTGTGAGTCTCGCGCTCGCGCTCGCGCTCGCGCTCGCGCTCGCGCTCGCT 78

Qy 21 ProTrpProGlyLysLeuValAsnProProLysAspLeuLysProArgGlyLysLys 40

Db 79 CCTGGCCAGGAAGATTGTTAATCCACCAAGGACTTAAAGAAACACCGTGGAAAGAA 138

Qy 41 CysPhePheValLysPheGlyThrGluAspHisAlaTrpLysValGluGlnLeu 60

Db 139 TGTCTCTTGTGAAGTT-TTTGGAACAGAGATCATGCTGGATCAAAGTGGAAACAGCTA 197

Qy 61 LysProTrpHisAlaHisLysGluMetIleLysLysLysLysLysLysLysLysLys 80

Db 198 AAGCTTTACCATGCTCAAGGAGGAGATGATTAAGATTAAAGGCTTAAACGGTTCAG 257

Qy 81 GluAlaValAspAlaValGluGluPheLeuArgArgLysLysLysLysLysLysLys 100

Db 258 CAAGCTGTGGATGCTGTGTGGAGAGTTCTTAAGGAGAGCCAAAGGAAAGACACATCA 317

Qy 101 SerHisAsnSerSerAspLysAsnArgArgAsnSerSerGluGluArgSerArgPro 120

Db 318 TCCACACTTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 377

Qy 121 AsnSerGlyAspGluLysArgLysLeuSerLeuSerLeuSerGluGlyLysValLys 140

Db 378 AACTCAGTGTGATGAGAAACCAAGCTTAGCTGTCTGAAGGAGGAGTGAAGAACATG 437

Qy 141 GlyGluGlyLysLysArgValSerSerGlySerSerGluArgGlySerLysSerPro 160

Db 438 GGAGAGGAAGGAAGAGGGTGAAGTTCAGGCTCTGCAGACAGAGGCTCCAAATGC---CTT 494

```
Qy 161 LysArgAlaGlnGluGlnSerProArgLysArgGlyArgProProLysAspGluLysAsp 180
Db |||||
495 AAAAGAGCCCAAGAGCAAGTCCCGGAGCGGGGTCCGCCCCCAAGAGATGAGAGGAC 554
Qy 181 LeuThrIleProGluSerSerThrValLysGlyMetMetAlaGlyProMetAlaAlaPhe 200
Db |||||
555 CTCACCATCCCTGAGTCTAGCACTGTAAGGGGATGATGGCTGGACCGCATGGCTGCATT 614
Qy 201 LysTrpGlnProThrAlaSerGluProValLysAspAlaAspProHisPheHisPhe 220
Db |||||
615 AAATGGCAGCCCAACAGCAGCCAGCCAGTCAAGAGATGCAGATCCTCATTTTCATCATTTT 674
Qy 221 LeuLeuSerGlnThrGluLysProAlaValCysTrpGlnAlaIleThrLysLysLeuLys 240
Db |||||
675 CTGTTGACCCAAACAGAGAGCCAGCTGTCTGTACCAAGGCATCACAAAGAGTTGAAA 734
Qy 241 IleCysGluApp 244
Db |||||
735 ATATGTGAAGAG 746
```

Search completed: December 8, 2003, 13:25:41
Job time : 2159 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 8, 2003, 13:25:45 ; Search time 304 Seconds
(without alignments)
8166.870 Million cell updates/sec

Title: US-09-987-755-1_COPY_34_780

Perfect score: 747
Sequence: 1 atggcggtgtgagctgctgcg.....aagacctcttcttcttagg 747

Scoring table: IDENTITY NUC
Gapop 10_0 , Gapext 1.0

Searched: 2201672 seqs, 1661799599 residues

Total number of hits satisfying chosen parameters: 4403344

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications_NA.*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	747	100.0	3202	11	US-09-987-755-1 Sequence 1, Appli
2	747	100.0	3256	15	US-10-103-313-190 Sequence 190, Appl
3	745	99.7	3172	15	US-10-103-313-33 Sequence 33, Appl
4	731	97.9	3727	13	US-10-067-482-1 Sequence 1, Appli
5	722.4	96.7	3707	15	US-10-103-313-133 Sequence 133, Appl
6	719.6	96.3	810	15	US-10-103-313-256 Sequence 256, Appl
7	53.8	7.2	865	15	US-10-198-846-6497 Sequence 6497, Ap
8	53.8	7.2	872	15	US-10-198-846-6652 Sequence 6652, Ap
9	53.8	7.2	890	15	US-10-198-846-9404 Sequence 9404, Ap
10	53.8	7.2	897	15	US-10-198-846-6712 Sequence 6712, Ap
11	53.8	7.2	901	15	US-10-198-846-6624 Sequence 6624, Ap
12	53.8	7.2	942	15	US-10-198-846-6578 Sequence 6578, Ap
13	53.8	7.2	1121	15	US-10-198-846-8409 Sequence 8409, Ap
14	53.2	7.1	822	15	US-10-198-846-9427 Sequence 9427, Ap
15	53.2	7.1	863	15	US-10-198-846-6557 Sequence 6557, Ap

16	52.8	7.1	842	15	US-10-198-846-4084	Sequence 4084, Ap
17	52.8	7.1	1127	15	US-10-198-846-8540	Sequence 8540, Ap
18	51.6	6.9	1038	15	US-10-198-846-9410	Sequence 9410, Ap
19	47	6.3	2920	11	US-09-866-050A-541	Sequence 541, App
20	47	6.3	2920	15	US-10-152-661-541	Sequence 541, App
21	44	5.9	1898	9	US-09-768-826-22	Sequence 22, Appli
22	44	5.9	3805	13	US-10-247-671-8	Sequence 8, Appli
23	41	5.5	288	10	US-09-960-352-1025	Sequence 1025, Ap
24	40.4	5.4	594	13	US-10-140-472-10	Sequence 10, Appl
25	40.4	5.4	594	13	US-10-141-761-10	Sequence 10, Appl
26	40.4	5.4	594	13	US-10-142-885-10	Sequence 10, Appl
27	40.4	5.4	594	13	US-10-158-790-10	Sequence 10, Appl
28	40.4	5.4	594	13	US-10-137-871-10	Sequence 10, Appl
29	40.4	5.4	594	13	US-10-140-805-10	Sequence 10, Appl
30	40.4	5.4	594	13	US-10-140-864-10	Sequence 10, Appl
31	40.4	5.4	594	13	US-10-140-923-10	Sequence 10, Appl
32	40.4	5.4	594	13	US-10-141-756-10	Sequence 10, Appl
33	40.4	5.4	594	13	US-10-141-759-10	Sequence 10, Appl
34	40.4	5.4	594	15	US-10-123-155-10	Sequence 10, Appl
35	40.4	5.4	594	16	US-10-146-731-10	Sequence 10, Appl
36	37.2	5.0	491	10	US-09-796-692-8648	Sequence 8648, Ap
37	37.2	5.0	491	15	US-10-040-862-8648	Sequence 145, App
38	37.2	5.0	1919	9	US-09-925-301-145	Sequence 83, Appl
39	37.2	5.0	4866	13	US-10-007-926A-83	Sequence 140, App
40	37.2	5.0	4866	15	US-10-171-581-140	Sequence 6607, Ap
41	36.4	4.9	552	15	US-10-198-846-6607	Sequence 24, Appl
42	36	4.8	1297	9	US-09-768-826-24	Sequence 23, Appl
43	36	4.8	2031	11	US-09-823-187-23	Sequence 307, App
44	36	4.8	2272	11	US-09-946-374-307	Sequence 307, App
45	36	4.8	2272	13	US-10-015-387A-307	Sequence 307, App

ALIGNMENTS

RESULT 1

US-09-987-755-1
; Sequence 1, Application US/09987755
; Publication No. US20030022312A1
; GENERAL INFORMATION:
; APPLICANT: Kunsch et al
; TITLE OF INVENTION: Human Hepatoma-Derived Growth Factor-2
; FILE REFERENCE: PF198D1C1
; CURRENT APPLICATION NUMBER: US/09/987,755
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: 09/263,625
; PRIOR FILING DATE: 1999-03-05
; PRIOR APPLICATION NUMBER: 08/464,600
; PRIOR FILING DATE: 1995-06-05
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 3202
; TYPE: DNA
; ORGANISM: human
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (34)..(780)
; OTHER INFORMATION:
US-09-987-755-1

Query Match	100.0%	Score 747;	DB 11;	Length 3202;
Best Local Similarity	100.0%	Pred. No. 3.9e-227;	Mismatches 0;	Indels 0;
Matches 747;	Conservative 0;			Gaps 0;
QY	1	ATGGCGGCTGTGAGTCTGCGCTCGCGGACTTGGTGTGGGGAACCTCGCGGATATCTCT	60	
Db	34	ATGGCGGCTGTGAGTCTGCGCTCGCGGACTTGGTGTGGGGAACCTCGCGGATATCTCT	93	
QY	61	CCTTGGCCAGGAAAGATTGTTAATCCAAAGGACTTGAAGAAACCTCGCGGAAGAA	120	
Db	94	CCTTGGCCAGGAAAGATTGTTAATCCAAAGGACTTGAAGAAACCTCGCGGAAGAA	153	

```
Qy 121 TGCTTCTTTGTGAAATTTTGGAAACAGAGATCATGCTGGATCAAAGTGGACAGCTG 180
Db 154 TGCTTCTTTGTGAAATTTTGGAAACAGAGATCATGCTGGATCAAAGTGGACAGCTG 213
Qy 181 AAGCCATATCATGCTCATAAAGAGGAATGATATAAAATTAACAAGGGTAAACGATTCCAG 240
Db 214 AAGCCATATCATGCTCATAAAGAGGAATGATATAAAATTAACAAGGGTAAACGATTCCAG 273
Qy 241 CAAGCGTAGATGCTGTCGAGAGTTCCTCAGGAGGCCAAAGGGAAGACCGAGCTCA 300
Db 274 CAAGCGTAGATGCTGTCGAGAGTTCCTCAGGAGGCCAAAGGGAAGACCGAGCTCA 333
Qy 301 TCCACAAATTTCTTGATGACAAATCGACGTAATTCAGTTCAGAGAGGCTCCAAATCCCCTCTG 360
Db 334 TCCACAAATTTCTTGATGACAAATCGACGTAATTCAGTTCAGAGAGGCTCCAAATCCCCTCTG 393
Qy 361 AACTCAGGTGATGAGAGCGCAACTTAGCTGTCTGAAGGGAGGTGAAGAGAACATG 420
Db 394 AACTCAGGTGATGAGAGCGCAACTTAGCTGTCTGAAGGGAGGTGAAGAGAACATG 453
Qy 421 GGAGAGGAAAGAGAGGGTGTCTTCAGGCTCTTCAGAGAGGCTCCAAATCCCCTCTG 480
Db 454 GGAGAGGAAAGAGAGGGTGTCTTCAGGCTCTTCAGAGAGGCTCCAAATCCCCTCTG 513
Qy 481 AAAAGAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTCCGCCCAAGAGGTGAAGAGAACATG 540
Db 514 AAAAGAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTCCGCCCAAGAGGTGAAGAGAACATG 573
Qy 541 CTCACCATCCCGAGTCTAGTACCGTGAAGGGAGTATGCCCGGACCGATGGCCGGGTTT 600
Db 574 CTCACCATCCCGAGTCTAGTACCGTGAAGGGAGTATGCCCGGACCGATGGCCGGGTTT 633
Qy 601 AAATGGCAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTCCGCCCAAGAGGTGAAGAGAACATG 660
Db 634 AAATGGCAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTCCGCCCAAGAGGTGAAGAGAACATG 693
Qy 661 CTGCTAAGCCAAACAGAGAGCCAGCTGTCTGTTTACAGGCAATCAGAGAGAGTTGAAA 720
Db 694 CTGCTAAGCCAAACAGAGAGCCAGCTGTCTGTTTACAGGCAATCAGAGAGAGTTGAAA 753
Qy 721 ATATGTGAAGACCTCTCTTCTCTAGG 747
Db 754 ATATGTGAAGACCTCTCTTCTCTAGG 780
```

RESULT 2

```
US-10-103-313-190
; Sequence 190, Application US/10103313
; Publication No. US20030082758A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PZ07C1
; CURRENT APPLICATION NUMBER: US/10/103,313
; CURRENT FILING DATE: 2002-03-12
; NUMBER OF SEQ ID NOS: 653
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 190
; LENGTH: 3256
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (3220)
; OTHER INFORMATION: n equals a.t.g. or c
US-10-103-313-190
```

```
Query Match 100.0%; Score 747; DB 15; Length 3256;
Best Local Similarity 100.0%; Pred. No. 4e-227;
Matches 747; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATGCGGCTGTGAGTCTGCGGCTCGGCGACTTGGTGTGGGGAAACTCGGCCGATATCCT 60
```

```
Db 34 ATGCGGCTGTGAGTCTGCGGCTCGGCGACTTGGTGTGGGGAAACTCGGCCGATATCCT 93
Qy 61 CTTTGGCCAGGAAGATTTGTAATCCACCAAGGACTTCAAGAAACCTCGCGAAGAAA 120
Db 94 CTTTGGCCAGGAAGATTTGTAATCCACCAAGGACTTCAAGAAACCTCGCGAAGAAA 153
Qy 121 TGTCTTCTTTGTGAAATTTTGGAAACAGAGATCATGCTGGATCAAAGTGGAAACAGCTG 180
Db 154 TGTCTTCTTTGTGAAATTTTGGAAACAGAGATCATGCTGGATCAAAGTGGAAACAGCTG 213
Qy 181 AAGCCATATCATGCTCATAAAGAGGAATGATATAAAATTAACAAGGGTAAACGATTCCAG 240
Db 214 AAGCCATATCATGCTCATAAAGAGGAATGATATAAAATTAACAAGGGTAAACGATTCCAG 273
Qy 241 CAAGCGTAGATGCTGTCGAGAGTTCCTCAGGAGGCCAAAGGGAAGACCGAGCTCA 300
Db 274 CAAGCGTAGATGCTGTCGAGAGTTCCTCAGGAGGCCAAAGGGAAGACCGAGCTCA 333
Qy 301 TCCACAAATTTCTTGATGACAAATCGACGTAATTCAGTTCAGAGAGGCTCCAAATCCCCTCTG 360
Db 334 TCCACAAATTTCTTGATGACAAATCGACGTAATTCAGTTCAGAGAGGCTCCAAATCCCCTCTG 393
Qy 361 AACTCAGGTGATGAGAGCGCAACTTAGCTGTCTGAAGGGAGGTGAAGAGAACATG 420
Db 394 AACTCAGGTGATGAGAGCGCAACTTAGCTGTCTGAAGGGAGGTGAAGAGAACATG 453
Qy 421 GGAGAGGAAAGAGAGGGTGTCTTCAGGCTCTTCAGAGAGGCTCCAAATCCCCTCTG 480
Db 454 GGAGAGGAAAGAGAGGGTGTCTTCAGGCTCTTCAGAGAGGCTCCAAATCCCCTCTG 513
Qy 481 AAAAGAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTCCGCCCAAGAGGTGAAGAGAACATG 540
Db 514 AAAAGAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTCCGCCCAAGAGGTGAAGAGAACATG 573
Qy 541 CTCACCATCCCGAGTCTAGTACCGTGAAGGGAGTATGCCCGGACCGATGGCCGGGTTT 600
Db 574 CTCACCATCCCGAGTCTAGTACCGTGAAGGGAGTATGCCCGGACCGATGGCCGGGTTT 633
Qy 601 AAATGGCAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTCCGCCCAAGAGGTGAAGAGAACATG 660
Db 634 AAATGGCAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTCCGCCCAAGAGGTGAAGAGAACATG 693
Qy 661 CTGCTAAGCCAAACAGAGAGCCAGCTGTCTGTTTACAGGCAATCAGAGAGAGTTGAAA 720
Db 694 CTGCTAAGCCAAACAGAGAGCCAGCTGTCTGTTTACAGGCAATCAGAGAGAGTTGAAA 753
Qy 721 ATATGTGAAGACCTCTCTTCTCTAGG 747
Db 754 ATATGTGAAGACCTCTCTTCTCTAGG 780
```

RESULT 3

```
US-10-103-313-33
; Sequence 33, Application US/10103313
; Publication No. US20030082758A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PZ07C1
; CURRENT APPLICATION NUMBER: US/10/103,313
; CURRENT FILING DATE: 2002-03-12
; NUMBER OF SEQ ID NOS: 653
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 33
; LENGTH: 3172
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-103-313-33
```

```
Query Match 99.7%; Score 745; DB 15; Length 3172;
Best Local Similarity 100.0%; Pred. No. 1.7e-226;
```

```
Matches 745; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 GCGGCTGTGAGTCTGCGGCTCGGCACTTGGTGTGGGGGAAACCTCGGCCGATATCTCC 62
Db 8 GCGGCTGTGAGTCTGCGGCTCGGCACTTGGTGTGGGGGAAACCTCGGCCGATATCTCC 67
QY 63 TTGGCCAGGAAAGATTGTTAATCCAAAGAGCTTGAAGAAACCTCGCGGAAAGAAATG 122
Db 68 TTGGCCAGGAAAGATTGTTAATCCAAAGAGCTTGAAGAAACCTCGCGGAAAGAAATG 127
QY 123 CTTCTTTTGTAAATTTTGGACAGAGATCATGCTGGATCAAGTGGACAGCTGAA 182
Db 128 CTTCTTTTGTAAATTTTGGACAGAGATCATGCTGGATCAAGTGGACAGCTGAA 187
QY 183 GCATATCATGCTCATAAGAGGAAATGATAAAATTAACAGGGTAAACGATTCCAGCA 242
Db 188 GCATATCATGCTCATAAGAGGAAATGATAAAATTAACAGGGTAAACGATTCCAGCA 247
QY 243 AGCGGTAGATGCTGCGAAGTTCCTCAGGAGAGCCAAAGGAAAGACCGACGTCATC 302
Db 248 AGCGGTAGATGCTGCGAAGTTCCTCAGGAGAGCCAAAGGAAAGACCGACGTCATC 307
QY 303 CCACAAATCTTCTGATGACAAATCGAGCTTAATTCAGTGGAGAGAGTAGGCCAAA 362
Db 308 CCACAAATCTTCTGATGACAAATCGAGCTTAATTCAGTGGAGAGAGTAGGCCAAA 367
QY 363 CTCAGGTGATGAGAGCGCAAACTTAGCTGTCTGAAGGGAAGGTGAAGAAACATGGG 422
Db 368 CTCAGGTGATGAGAGCGCAAACTTAGCTGTCTGAAGGGAAGGTGAAGAAACATGGG 427
QY 423 AGAAGGAAAGAGAGGGTGTCTTCAGGCTTTTCAGAGAGAGGCTCCAAATCCCTCTGAA 482
Db 428 AGAAGGAAAGAGAGGGTGTCTTCAGGCTTTTCAGAGAGAGGCTCCAAATCCCTCTGAA 487
QY 483 AAGAGCCCAAGAGCAAAAGTCCCGAGAGCGGGTGGCCCCCAAGAGATGAGAGATCT 542
Db 488 AAGAGCCCAAGAGCAAAAGTCCCGAGAGCGGGTGGCCCCCAAGAGATGAGAGATCT 547
QY 543 CACCATCCCGGAGTCTAGTACCGTGAAGGGGATGATGGCCGACCGATGGCCGCTTTAA 602
Db 548 CACCATCCCGGAGTCTAGTACCGTGAAGGGGATGATGGCCGACCGATGGCCGCTTTAA 607
QY 603 ATGGCAGCCCAACCGCAAGCGAGCTGTTAAAGATGAGATCTCATTTTCATCTTCT 662
Db 608 ATGGCAGCCCAACCGCAAGCGAGCTGTTAAAGATGAGATCTCATTTTCATCTTCT 667
QY 663 GCTAAGCCAAACAGAGAGCGAGCTGTTGTTACAGGCAATCAGAGAGAGTTGAAAT 722
Db 668 GCTAAGCCAAACAGAGAGCGAGCTGTTGTTACAGGCAATCAGAGAGAGTTGAAAT 727
QY 723 ATGTGAAGACCTCTTCTTCTTAGG 747
Db 728 ATGTGAAGACCTCTTCTTCTTAGG 752
```

```
RESULT 4
US-10-067-482-1
; Sequence 1, Application US/10067482
; Publication No. US20030148407A1
; GENERAL INFORMATION:
; APPLICANT: OxiGene Technologies, Inc.
; TITLE OF INVENTION: Angiogenesis Dehydrogenase Gene
; FILE REFERENCE: 1U 102 R1
; CURRENT APPLICATION NUMBER: US/10/067,482
; CURRENT FILING DATE: 2002-02-07
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 3727
; TYPE: DNA
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: CDS
```

```
; LOCATION: (28)..(1689)
; OTHER INFORMATION:
US-10-067-482-1
Query Match 97.9%; Score 731; DB 13; Length 3727;
Best Local Similarity 100.0%; Pred. No. 5.4e-222; Indels 0; Gaps 0;
Matches 731; Conservative 0; Mismatches 0;
QY 1 ATGGCGGCTGTGAGTCTGCGGCTCGGCACTTGGTGTGGGGGAAACCTCGGCCGATATCT 60
Db 28 ATGGCGGCTGTGAGTCTGCGGCTCGGCACTTGGTGTGGGGGAAACCTCGGCCGATATCT 87
QY 61 CTTTGGCCAGGAAAGATTGTTAATCCAAAGAGCTTGAAGAAACCTCGCGGAAAGAA 120
Db 88 CTTTGGCCAGGAAAGATTGTTAATCCAAAGAGCTTGAAGAAACCTCGCGGAAAGAA 147
QY 121 TGTCTTTTGTAAATTTTGGAAACAGAGATCATGCTCGATCAAAAGTGGAAACAGCTG 180
Db 148 TGTCTTTTGTAAATTTTGGAAACAGAGATCATGCTCGATCAAAAGTGGAAACAGCTG 207
QY 181 AAGCCATATCATGCTCATAAAGAGGAAATGATAAAATTAACAGGGTAAACGATTCCAG 240
Db 208 AAGCCATATCATGCTCATAAAGAGGAAATGATAAAATTAACAGGGTAAACGATTCCAG 267
QY 241 CAAGCGGTAGATGCTGTGCGAAGATTCTCAGGAGAGCCAAAGGAAAGACCGACGTC 300
Db 268 CAAGCGGTAGATGCTGTGCGAAGATTCTCAGGAGAGCCAAAGGAAAGACCGACGTC 327
QY 301 TCCACAAATCTTCTGATGACAAATCGAGCTTAATTCAGTGGAGAGAGTAGGCCCA 360
Db 328 TCCACAAATCTTCTGATGACAAATCGAGCTTAATTCAGTGGAGAGAGTAGGCCCA 387
QY 361 AACTCAGGTGATGAGAGCGCAAACTTAGCTGTCTGAAGGGAAGGTGAAGAAACATG 420
Db 388 AACTCAGGTGATGAGAGCGCAAACTTAGCTGTCTGAAGGGAAGGTGAAGAAACATG 447
QY 421 GGAGAGGAAAGAGAGGGTGTCTTCAGGCTTTTCAGAGAGAGGCTCCAAATCCCTCTG 480
Db 448 GGAGAGGAAAGAGAGGGTGTCTTCAGGCTTTTCAGAGAGAGGCTCCAAATCCCTCTG 507
QY 481 AAAGAGCCCAAGAGCAAAAGTCCCGAGAGCGGGTGGCCCCCAAGAGATGAGAGAT 540
Db 508 AAAGAGCCCAAGAGCAAAAGTCCCGAGAGCGGGTGGCCCCCAAGAGATGAGAGAT 567
QY 541 CTCACCATCCCGAGTCTAGTACCGTGAAGGGGATGATGGCCGACCGATGGCCGCTTT 600
Db 568 CTCACCATCCCGAGTCTAGTACCGTGAAGGGGATGATGGCCGACCGATGGCCGCTTT 627
QY 601 AAATGGCAGCCCAACCGCAAGCGAGCTGTTAAAGATGAGATCTCATTTTCATCTTTC 660
Db 628 AAATGGCAGCCCAACCGCAAGCGAGCTGTTAAAGATGAGATCTCATTTTCATCTTTC 687
QY 661 CTGCTTAAGCCAAACAGAGAGCGAGCTGTTGTTACAGGCAATCAGAGAGAGTTGAA 720
Db 688 CTGCTTAAGCCAAACAGAGAGCGAGCTGTTGTTACAGGCAATCAGAGAGAGTTGAA 747
QY 721 ATATGTGAAGA 731
Db 748 ATATGTGAAGA 758
```

```
RESULT 5
US-10-103-313-133
; Sequence 133, Application US/10103313
; Publication No. US20030082758A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJ207C1
; CURRENT APPLICATION NUMBER: US/10/103,313
; CURRENT FILING DATE: 2002-03-12
; NUMBER OF SEQ ID NOS: 653
; Prior Application removed - See File Wrapper or Palm
```

```
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 133
; LENGTH: 3707
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-103-313-133

Query Match      96.7%; Score 722.4; DB 15; Length 3707;
Best Local Similarity 99.9%; Pred. No. 3e-219;
Matches 723; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 8 CTGTGAGTCTCGGCTCGGCGACTTGGTGTGGGGAAACTCGGCCGATATCCTCTTGGC 67
Db |||
Qy 18 CGGTGAGTCTCGGCTCGGCGACTTGGTGTGGGGAAACTCGGCCGATATCCTCTTGGC 77
Db |||
Qy 68 CAGAAAGATGTTTAATCCACCAAGGACTTGAAGAACTCGGGGAAAGAAATGCTTCT 127
Db |||
Qy 78 CAGAAAGATGTTTAATCCACCAAGGACTTGAAGAACTCGGGGAAAGAAATGCTTCT 137
Db |||
Qy 128 TTGTGAAATTTTGGAAACAGAGATCATGCTGGATCAAGTGGAAACAGCTGAAGCAT 187
Db |||
Qy 138 TTGTGAAATTTTGGAAACAGAGATCATGCTGGATCAAGTGGAAACAGCTGAAGCAT 197
Db |||
Qy 188 ATCATGCTCTAAGAGGAAATGATAAAATTAACAAAGGTTAAACGATTCAGCAAGCGG 247
Db |||
Qy 198 ATCATGCTCTAAGAGGAAATGATAAAATTAACAAAGGTTAAACGATTCAGCAAGCGG 257
Db |||
Qy 248 TAGATGCTGCGAAGAGTTCTCAGGAGAGCCAAAGGAAAGACAGACGTCATCCACCA 307
Db |||
Qy 258 TAGATGCTGCGAAGAGTTCTCAGGAGAGCCAAAGGAAAGACAGACGTCATCCACCA 317
Db |||
Qy 308 ATTCTTCTGATGACAAGATCGAGTAATTCAGTGGAGGAGAGAGTAGGCCAACTCAG 367
Db |||
Qy 318 ATTCTTCTGATGACAAGATCGAGTAATTCAGTGGAGGAGAGAGTAGGCCAACTCAG 377
Db |||
Qy 368 GTGATGAGAGCGCAAACTTAGCCTGTCTGAAGGGAAGGTGAAGAAAGATCGGGAAG 427
Db |||
Qy 378 GTGATGAGAGCGCAAACTTAGCCTGTCTGAAGGGAAGGTGAAGAAAGATCGGGAAG 437
Db |||
Qy 428 GAAGAAAGAGGGTGTCTTACGGCTTTCAGAGAGAGCTCCAAATCCCTCTGAAAGAG 487
Db |||
Qy 438 GAAAGAAAGAGGGTGTCTTACGGCTTTCAGAGAGAGCTCCAAATCCCTCTGAAAGAG 497
Db |||
Qy 488 CCAAGAGCAAGTCCCGGAGCGGGTCCGCCCCCAAGAGGATGAAAGGATCTACCA 547
Db |||
Qy 498 CCAAGAGCAAGTCCCGGAGCGGGTCCGCCCCCAAGAGGATGAAAGGATCTACCA 557
Db |||
Qy 548 TCCCGGAGTCTAGTACCGTGAAGGGGATGATGGCCGACCGATGGCCGCGCTTTAAATGGC 607
Db |||
Qy 558 TCCCGGAGTCTAGTACCGTGAAGGGGATGATGGCCGACCGATGGCCGCGCTTTAAATGGC 617
Db |||
Qy 608 AGCCAAACCGAAGCGAGCTGTTAAAGATGAGATCCTCAATTCATCATTTCTGCTAA 667
Db |||
Qy 618 AGCCAAACCGAAGCGAGCTGTTAAAGATGAGATCCTCAATTTCCATCATTTCTGCTAA 677
Db |||
Qy 668 GCCAAACAGAGAGCCAGCTGTTTACCGAGCAATCAAGAGAGTTGAAATATGTG 727
Db |||
Qy 678 GCCAAACAGAGAGCCAGCTGTTTACCGAGCAATCAAGAGAGTTGAAATATGTG 737
Db |||
Qy 728 AAGA 731
Db |||
Qy 738 AAGA 741
Db |||
```

RESULT 6

```
US-10-103-313-256
; Sequence 256, Application US/10103313
; Publication No. US20030082758A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P0207C1
; CURRENT APPLICATION NUMBER: US/10/103.313
```

```
; CURRENT FILING DATE: 2002-03-12
; NUMBER OF SEQ ID NOS: 653
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 256
; LENGTH: 810
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (12)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (43)
; OTHER INFORMATION: n equals a,t,g, or c
; US-10-103-313-256

Query Match      96.3%; Score 719.6; DB 15; Length 810;
Best Local Similarity 99.6%; Pred. No. 1e-218;
Matches 719; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 10 GTGAGTCTGGGCTCGGCGACTTGGTGTGGGGAAACTCGGCCGATATCCTCTTGGCCA 69
Db |||
Qy 2 GTGAGTCTGCGCTCGGCGACTTGGTGTGGGGAAACTCGGCCGATATCCTCTTGGCCA 61
Db |||
Qy 70 GGAAGATGTTTAATCCACCAAGGACTTGAAGAACTCGCGGAAAGAAATGCTTCTTT 129
Db |||
Qy 62 GGAAGATGTTTAATCCACCAAGGACTTGAAGAACTCGCGGAAAGAAATGCTTCTTT 121
Db |||
Qy 130 GTGAAATTTTTTGGAAACAGAGATCATGCTTGAAGTCAAAAGTGAACAGCTGAAGCCATAT 189
Db |||
Qy 122 GTGAAATTTTTTGGAAACAGAGATCATGCTTGAAGTCAAAAGTGAACAGCTGAAGCCATAT 181
Db |||
Qy 190 CATGCTCTAAGAGGAAATGATAAAATTAACAAAGGTTAAACGATTCAGCAAGCGGTA 249
Db |||
Qy 182 CATGCTCTAAGAGGAAATGATAAAATTAACAAAGGTTAAACGATTCAGCAAGCGGTA 241
Db |||
Qy 250 GATGCTGTGGAAGAGTTCTCAGGAGAGCCAAAGGAAAGACAGACGTCATCCCAAT 309
Db |||
Qy 242 GATGCTGTGGAAGAGTTCTCAGGAGAGCCAAAGGAAAGACAGACGTCATCCCAAT 301
Db |||
Qy 310 TCTTCTGATGACAAGAAATCGAATAATTCAGTGGAGAGAGAAAGTAGGCCAAATCAGGT 369
Db |||
Qy 302 TCTTCTGATGACAAGAAATCGAATAATTCAGTGGAGAGAGAAAGTAGGCCAAATCAGGT 361
Db |||
Qy 370 GATGAGAGCGCAAACTTAGCCTGTCTGAAGGGAAGTGAAGAAACATCGGAGAGGA 429
Db |||
Qy 362 GATGAGAGCGCAAACTTAGCCTGTCTGAAGGGAAGTGAAGAAACATCGGAGAGGA 421
Db |||
Qy 430 AAGAAGAGGTTGTCTTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTGAAAGAGCC 489
Db |||
Qy 422 AAGAAGAGGTTGTCTTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTGAAAGAGCC 481
Db |||
Qy 490 CAAGACAAAGTCCCGGAAAGCGGGTCCGCCCCCAAGGATGAGAAGGATCTCACCATC 549
Db |||
Qy 482 CAAGACAAAGTCCCGGAAAGCGGGTCCGCCCCCAAGGATGAGAAGGATCTCACCATC 541
Db |||
Qy 550 CCGGAGTCTAGTACCGTGAAGGGGATGATGGCCGACCGATGGCCGCGCTTTAAATGGCAG 609
Db |||
Qy 542 CCGGAGTCTAGTACCGTGAAGGGGATGATGGCCGACCGATGGCCGCGCTTTAAATGGCAG 601
Db |||
Qy 610 CCAACCGCAAGCGAGCTGTTAAAGATGAGATCCCTCATTTCCATCATTTCTGCTAAGC 669
Db |||
Qy 602 CCAACCGCAAGCGAGCTGTTAAAGATGAGATCCCTCATTTCCATCATTTCTGCTAAGC 661
Db |||
Qy 670 CAAACAGAGAGCGAGCTGTTTACCGAGCAATCAAGAGAGTTGAAATATGTGAA 729
Db |||
Qy 662 CAAACAGAGAGCGAGCTGTTTACCGAGCAATCAAGAGAGTTGAAATATGTGAA 721
Db |||
Qy 730 GA 731
Db |||
Qy 722 GA 723
Db |||
```

```
RESULT 7
US-10-198-846-6497
; Sequence 6497, Application US/10198846
; Publication No. US20030099974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6497
; LENGTH: 865
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 262, 406, 416, 417, 437, 449, 451, 455, 465, 474, 494, 498,
; LOCATION: 515, 517, 550, 584, 587, 607, 619, 620, 626, 652, 662, 663,
; LOCATION: 666, 691, 697, 711, 722, 728, 764, 776, 777, 802, 804, 810,
; LOCATION: 812, 819, 822, 824, 825, 826, 829, 831, 836, 848, 849
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 850, 856, 860, 862, 864
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-6497

Query Match      7.2%; Score 53.8; DB 15; Length 865;
Best Local Similarity 89.2%; Pred. No. 2.7e-06;
Matches 58; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ATGGCGGCTGTGAGTCTCGCGCTCGCGACTTGGTGTGGGGAAACTCGCGCATATCCT 60
Db 67 ATGGCGGCTGTGAGTCTCGCGCTCGCGACTTGGTGTGGGGAAACTCGCGCATATCCT 126

QY 61 CCTTG 65
Db 127 ACTAG 131

RESULT 8
US-10-198-846-6652
; Sequence 6652, Application US/10198846
; Publication No. US20030099974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6652
; LENGTH: 872
; TYPE: DNA
; ORGANISM: Homo sapiens

Query Match      7.2%; Score 53.8; DB 15; Length 865;
Best Local Similarity 89.2%; Pred. No. 2.7e-06;
Matches 58; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ATGGCGGCTGTGAGTCTCGCGCTCGCGACTTGGTGTGGGGAAACTCGCGCATATCCT 60
Db 67 ATGGCGGCTGTGAGTCTCGCGCTCGCGACTTGGTGTGGGGAAACTCGCGCATATCCT 126

QY 61 CCTTG 65
Db 127 ACTAG 131

RESULT 9
US-10-198-846-9404
; Sequence 9404, Application US/10198846
; Publication No. US20030099974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9404
; LENGTH: 890
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 515, 565, 609, 617, 623, 633, 637, 648, 658, 659, 703, 709,
; LOCATION: 719, 723, 727, 742, 749, 753, 755, 757, 763, 767, 768, 770,
; LOCATION: 774, 778, 779, 781, 782, 783, 786, 790, 792, 799, 802, 806,
; LOCATION: 814, 815, 819, 825, 829, 836, 838, 849, 850, 858, 865
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 868, 873, 874, 877, 878, 879, 881, 885, 887, 889, 890
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-9404

Query Match      7.2%; Score 53.8; DB 15; Length 890;
Best Local Similarity 89.2%; Pred. No. 2.7e-06;
Matches 58; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ATGGCGGCTGTGAGTCTCGCGCTCGCGACTTGGTGTGGGGAAACTCGCGCATATCCT 60
Db 65 ATGGCGGCTGTGAGTCTCGCGCTCGCGACTTGGTGTGGGGAAACTCGCGCATATCCT 124

QY 61 CCTTG 65
Db 125 ACTAG 129

RESULT 10
US-10-198-846-6712
; Sequence 6712, Application US/10198846
; Publication No. US20030099974A1
```

```

; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6712
; LENGTH: 897
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 141, 178, 247, 310, 405, 423, 438, 440, 445, 446, 452, 453,
; LOCATION: 487, 490, 492, 503, 515, 517, 522, 531, 561, 563, 564, 587,
; LOCATION: 599, 604, 623, 643, 650, 676, 720, 721, 727, 729, 736, 744,
; LOCATION: 747, 749, 752, 755, 758, 764, 766, 770, 781, 786, 792
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 799, 800, 821, 826, 831, 836, 839, 843, 850, 854, 866, 870,
; LOCATION: 799, 800, 821, 826, 831, 836, 839, 843, 850, 854, 866, 870,
; LOCATION: 874, 879, 884, 888
; OTHER INFORMATION: n = A,T,C or G
; US-10-198-846-6712

Query Match 7.2%; Score 53.8; DB 15; Length 897;
Best Local Similarity 89.2%; Pred. No. 2.8e-06;
Matches 58; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 ATGGCGGCTGTGAGTCTGGCGGCTGGCGGAACTTGGTGGGGAACTCGGCCGATATCCT 60
Db 64 ATGGCGGCTGTGAGTCTGGCGGCTGGCGGAACTTGGTGGGGAACTCGGCCGCTCTAGA 123

Qy 61 CCTTG 65
Db 124 ACTAG 128

RESULT 11
US-10-198-846-6624
; Sequence 6624, Application US/10/198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6624
; LENGTH: 901
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 3, 4, 11, 13, 14, 33, 298, 299, 300, 301, 302, 303, 304,
; LOCATION: 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316,
; LOCATION: 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328,
; LOCATION: 329, 330, 331, 332, 334, 335, 336, 337, 338, 339, 340
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352,
; LOCATION: 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364,
; LOCATION: 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376,
; LOCATION: 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399,
; LOCATION: 400, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410, 411,
; LOCATION: 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423,
; LOCATION: 424, 425, 426, 427, 428, 429, 430, 431, 432, 433, 434
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 435, 436, 437, 438, 439, 440, 441, 442, 443, 444, 445, 446,
; LOCATION: 447, 448, 449, 450, 451, 452, 453, 454, 455, 456, 457, 458,
; LOCATION: 459, 460, 461, 462, 463, 464, 465, 466, 467, 468, 469, 470,
; LOCATION: 471, 472, 473, 474, 475, 476, 477, 478, 479, 480, 481
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493,
; LOCATION: 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504, 505,
; LOCATION: 506, 507, 508, 509, 510, 511, 512, 513, 514, 515, 516, 517,
; LOCATION: 518, 519, 520, 521, 522, 523, 524, 525, 526, 527, 528
; OTHER INFORMATION: n = A,T,C or G
; US-10-198-846-6624

Query Match 7.2%; Score 53.8; DB 15; Length 901;
Best Local Similarity 89.2%; Pred. No. 2.8e-06;
Matches 58; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 ATGGCGGCTGTGAGTCTGGCGGCTGGCGGAACTTGGTGGGGAACTCGGCCGATATCCT 60
Db 90 ATGGCGGCTGTGAGTCTGGCGGCTGGCGGAACTTGGTGGGGAACTCGGCCGCTCTAGA 149

Qy 61 CCTTG 65
Db 150 ACTAG 154

RESULT 12
US-10-198-846-6578
; Sequence 6578, Application US/10/198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6578
```

```
; LENGTH: 942
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 325, 393, 415, 458, 478, 506, 511, 518, 520, 525, 529, 531,
; LOCATION: 534, 539, 552, 567, 575, 593, 594, 596, 598, 600, 601, 602,
; LOCATION: 603, 608, 613, 618, 625, 631, 633, 635, 641, 643, 644, 645,
; LOCATION: 652, 653, 658, 660, 666, 669, 676, 693, 694, 699, 700
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 703, 705, 707, 714, 715, 719, 721, 722, 725, 726, 736, 737,
; LOCATION: 741, 753, 760, 762, 772, 773, 775, 783, 785, 804, 805,
; LOCATION: 807, 809, 816, 817, 822, 826, 835, 838, 847, 848, 849, 850,
; LOCATION: 853, 854, 855, 856, 857, 863, 868, 869, 872, 875, 876
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 883, 885, 888, 889, 897, 899, 901, 904, 905, 908, 912, 915,
; LOCATION: 921, 923, 930, 937, 940
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-6578

Query Match          7.2%; Score 53.8; DB 15; Length 942;
Best Local Similarity 89.2%; Pred. No. 2.8e-06;
Matches 58; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ATGGCGGCTGTGAGTCTCGGCTCGCGACTTGGTGTGGGGAACCTCGCGGATATCCT 60
DB 56 ATGGCGGCTGTGAGTCTCGGCTCGCGACTTGGTGTGGGGAACCTCGCGGCTCTAGA 115

QY 61 CCTTG 65
DB 116 ACTAG 120

RESULT 13
US-10-198-846-8409
; Sequence 8409, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; PRIOR FILING DATE: 2002-07-18
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8409
; LENGTH: 1121
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 3, 4, 6, 12, 13, 17, 22, 23, 27, 29, 32, 38, 39, 47, 48,
; LOCATION: 50, 51, 54, 62, 63, 66, 67, 70, 71, 72, 73, 75, 77, 79,
; LOCATION: 80, 81, 82, 84, 86, 87, 88, 90, 92, 93, 99, 100, 101, 110,
; LOCATION: 111, 297, 428, 481, 563, 685, 689, 736, 751, 756, 761, 764
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 774, 775, 781, 782, 792, 813, 848, 851, 857, 863, 868, 875,
; LOCATION: 882, 886, 911, 926, 927, 938, 940, 941, 942, 951, 952, 964,
; LOCATION: 966, 967, 971, 979, 985, 987, 997, 1004, 1032, 1034, 1050,
; LOCATION: 1062, 1064, 1092, 1097, 1099, 1107, 1114, 1115, 1116
```

```
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-8409

Query Match          7.2%; Score 53.8; DB 15; Length 1121;
Best Local Similarity 89.2%; Pred. No. 3.1e-06;
Matches 58; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ATGGCGGCTGTGAGTCTCGGCTCGCGACTTGGTGTGGGGAACCTCGCGGATATCCT 60
DB 178 ATGGCGGCTGTGAGTCTCGGCTCGCGACTTGGTGTGGGGAACCTCGCGGCTCTAGA 237

QY 61 CCTTG 65
DB 238 ACTAG 242

RESULT 14
US-10-198-846-9427/c
; Sequence 9427, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9427
; LENGTH: 822
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 636, 661, 665, 701, 713, 723, 729, 738, 744, 745, 748, 750,
; LOCATION: 762, 777, 778, 785, 792, 794, 817
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-9427

Query Match          7.1%; Score 53.2; DB 15; Length 822;
Best Local Similarity 94.8%; Pred. No. 4.1e-06;
Matches 55; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGGCGGCTGTGAGTCTCGGCTCGCGACTTGGTGTGGGGAACCTCGCGGATATC 58
DB 76 ATGGCGGCTGTGAGTCTCGGCTCGCGACTTGGTGTGGGGAACCTCGCGGCCACC 19

RESULT 15
US-10-198-846-6557/c
; Sequence 6557, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
```

```
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6557
; LENGTH: 863
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 562, 571, 601, 625, 629, 639, 667, 680, 709, 710, 718, 726,
; LOCATION: 746, 749, 758, 762, 766, 792, 794, 798, 803, 809, 813, 816,
; LOCATION: 823, 824, 829, 833, 837, 849, 850, 852, 854
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-6557

Query Match      7.1%; Score 53.2; DB 15; Length 863;
Best Local Similarity 94.8%; Pred. No. 4.2e-06;
Matches 55; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 ATGGCGGCTGTGAGTCTGCGGCTCGGCGACTTGGTGTGGGGAAACTCGGCCGATATC 58
      |||||
Db      78 ATGGCGGCTGTGAGTCTGCGGCTCGGCGACTTGGTGTGGGGAAACTCGGCCGCCACC 21
```

Search completed: December 8, 2003, 15:02:31
Job time : 304 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 8, 2003, 12:48:26 ; Search time 2082 Seconds
(without alignments)
8720.194 Million cell updates/sec

Title: US-09-987-755-1_COPY_34_780

Perfect score: 747

Sequence: 1 atggcggtgtgagctgcg.....aagacctcttctcttagg 747

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estnu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_est1:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_pbg:*

27: em_gss_vrl:*

28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	730.2	97.8	1201	13	EX331741
2	729.4	97.6	3741	11	BC032855
3	727	97.3	870	13	BQ720854
4	726.4	97.2	1039	12	BM808810

5	726	97.2	884	13	BU158924
6	718	96.1	1003	9	AL536040
7	700.4	93.8	1005	12	BM808969
8	692.8	92.7	755	12	BI551885
9	673	90.1	778	12	BI600148
10	668.8	89.5	816	12	BI459800
11	647	86.6	1138	12	BM906992
12	634.2	84.9	993	13	BU520128
13	633.4	84.8	899	13	BQ892009
14	633.2	84.8	890	13	BQ918182
15	630.2	84.4	3239	11	AK014456
16	629	84.2	815	12	BI767890
17	623.6	83.5	829	13	BU593961
18	623	83.4	3250	11	AK040841
19	622.2	83.3	929	13	BQ934274
20	614.2	82.2	705	13	BU708303
21	614.2	82.2	813	12	BI691175
22	591.4	79.2	704	13	BQ768940
23	589.2	78.9	998	10	BB611224
24	584.4	78.2	773	12	BI103792
25	580.2	77.7	685	10	BB617747
26	575.2	77.0	682	14	BY735423
27	567.2	75.9	823	12	BI150279
28	561.2	75.1	801	14	CA514495
29	553.6	74.1	660	14	BY753847
30	550.2	73.7	659	12	BM947061
31	547.4	73.3	646	14	BY745429
32	544.8	72.9	671	12	BI732831
33	544.2	72.9	896	10	BE538528
34	540.4	72.3	732	12	BI218947
35	518.2	69.4	786	12	BG976825
36	517.6	69.3	908	10	BG294400
37	515.8	69.0	775	14	CB601170
38	506.8	67.8	515	10	BE207054
39	503.8	67.4	532	9	AL702600
40	498.4	66.7	642	14	BY745902
41	497.2	66.6	595	14	CB584042
42	494.4	66.2	816	13	BU373180
43	491.6	65.8	821	12	BG870714
44	490	65.6	574	12	BM288574
45	482.8	64.6	632	14	BY743220

ALIGNMENTS

RESULT 1
BX331741
LOCUS
DEFINITION BX331741 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
CDNA clone CSODC001YL09 5-PRIME, mRNA sequence.
ACCESSION BX331741
VERSION BX331741.1 GI:30335078
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1201)
AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 7982.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CSODC001CRO5QBP1&cluster=7982.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600

FEATURES		Paraday Avenue Genoscope sequence ID : CS0DC001CF05QP1.	
source	1.	Location/Qualifiers	
		/organism="Homo sapiens"	
		/mol_type="mRNA"	
		/db_xref="taxon:9606"	
		/clone="CS0DC001YL09"	
BASE COUNT	349 a	261 c	332 g
	240 t	19 others	
ORIGIN			
Query Match			
Best Local Similarity 97.8%; Score 730.2; DB 13; Length 1201;			
Matches 730; Conservative 1; Mismatches 0; Indels 0; Gaps 0;			
QY	1	ATGGCGCTGTGAGTCTGGCGCTCGCGACTTGGTGTGGGGAACTCGGCCGATATCCT	60
DB	75	ATGGCGCTGTGAGTCTGGCGCTCGCGACTTGGTGTGGGGAACTCGGCCGATATCCT	134
QY	61	CCTTGGCCAGAAAGATTGTTAATCCACCAAGACTTGAAGAACTCGCGGAAGAA	120
DB	135	CCTTGGCCAGAAAGATTGTTAATCCACCAAGACTTGAAGAACTCGCGGAAGAA	194
QY	121	TGCTTTCTTTGAAATTTTGGAAACAGAAAGATCATGCTCGATCAAAAGTGGACACGCTG	180
DB	195	TGCTTTCTTTGAAATTTTGGAAACAGAAAGATCATGCTCGATCAAAAGTGGACACGCTG	254
QY	181	AAGCCATATCATGCTCATAAAGAGAAATGATAAAATTAACAGGGTAACCATTCCTCAG	240
DB	255	AAGCCATATCATGCTCATAAAGAGAAATGATAAAATTAACAGGGTAACCATTCCTCAG	314
QY	241	CAAGCGGTAGATGCTGTCGAAGAGTTCCTCAGAGAGCCAAAGGAAAGACAGAGCTCA	300
DB	315	CAAGCGGTAGATGCTGTCGAAGAGTTCCTCAGAGAGCCAAAGGAAAGACAGAGCTCA	374
QY	301	TCCACAAATTTCTTGATGACAAAGAAATCGACGTAATTCAGTGAGGAGAGAAAGTAGGCCA	360
DB	375	TCCACAAATTTCTTGATGACAAAGAAATCGACGTAATTCAGTGAGGAGAGAAAGTAGGCCA	434
QY	361	AATCAGGTGATGAGAGCCCAACTTAGCTGTCTGAAGGGAAGGTGAGAGAACATG	420
DB	435	AATCAGGTGATGAGAGCCCAACTTAGCTGTCTGAAGGGAAGGTGAGAGAACATG	494
QY	421	GGAGAAGGAAAGAGGGGTGCTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTG	480
DB	495	GGAGAAGGAAAGAGGGGTGCTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTG	554
QY	481	AAAAGAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTGGCCCCCAAGAGATGAGAAGGAT	540
DB	555	AAAAGAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTGGCCCCCAAGAGATGAGAAGGAT	614
QY	541	CTCACCATCCCGAGTCTAGTACGTTGAAGGGAATGATGCCCGGATGGCGGTTT	600
DB	615	CTCACCATCCCGAGTCTAGTACGTTGAAGGGAATGATGCCCGGATGGCGGTTT	674
QY	601	AAATGGCAGCAACCGCAAGCGCTGTTAAAGATGCAGATCCTCATTTTCATATTC	660
DB	675	AAATGGCAGCAACCGCAAGCGCTGTTAAAGATGCAGATCCTCATTTTCATATTC	734
QY	661	CTGCTAAGCCAAACAGAGAGCCAGTGTCTGTTACAGGCAATACAGAGAAAGTTGAA	720
DB	735	CTGCTAAGCCAAACAGAGAGCCAGTGTCTGTTACAGGCAATACAGAGAAAGTTGAA	794
QY	721	ATATGTGAACA	731
DB	795	ATATGTGAACA	805
RESULT 2			
BC032855			
LOCUS			
DEFINITION			
Homo sapiens, Similar to RIKEN CDNA 3930401K13 gene, clone			
IMAGE:5265561, mRNA.			
ACCESSION			
BC032855			
VERSION			
BC032855.1			
KEYWORDS			
HTC.			
SOURCE			
Homo sapiens (human)			
ORGANISM			
Homo sapiens			
REFERENCE			
1 (bases 1 to 3741)			
AUTHORS			
Strausberg, R.			
TITLE			
Direct Submission			
JOURNAL			
Submitted (07-JUN-2002) National Institutes of Health, Mammalian			
Gene Collection (MGC), Cancer Genomics Office, National Cancer			
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,			
USA			
REMARK			
COMMENT			
NIH-MGC Project URL: http://mgc.nci.nih.gov			
Contact: MGC help desk			
Email: cgabbs@mail.nih.gov			
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.			
cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki			
Toshiyuki and Piero Carninci (RIKEN)			
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)			
DNA Sequencing by: Institute for Systems Biology			
http://www.systemsbio.org			
Contact: amadan@systemsbiology.org			
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettman, Anuradha			
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting			
Clone distribution: MGC clone distribution information can be found			
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov			
Series: IRAK Plate: 47 Row: J Column: 23			
This clone was selected for full length sequencing because it			
passed the following selection criteria: Hexamer frequency ORF			
analysis, similarity but not identity to protein			
This clone has the following problem: retained intron.			
FEATURES			
source			
1..3741			
/organism="Homo sapiens"			
/mol_type="mRNA"			
/db_xref="taxon:9606"			
/clone="IMAGE:5265561"			
/tissue_type="Testis"			
/clone_lib="NIH MGC_97"			
/lab_host="DH10B"			
/note="Vector: pBluescript"			
BASE COUNT			
940 a 959 c 993 g 849 t			
ORIGIN			
Query Match			
Best Local Similarity 97.6%; Score 729.4; DB 11; Length 3741;			
Matches 730; Conservative 0; Mismatches 1; Indels 0; Gaps 0;			
QY	1	ATGGCGCTGTGAGTCTGGCGCTCGCGACTTGGTGTGGGGAACTCGGCCGATATCCT	60
DB	29	ATGGCGCTGTGAGTCTGGCGCTCGCGACTTGGTGTGGGGAACTCGGCCGATATCCT	88
QY	61	CCTTGGCCAGGAAAGATTGTTAATCCACCAAGGACTTGAAGAACTCGCGGAAGAA	120
DB	89	CCTTGGCCAGGAAAGATTGTTAATCCACCAAGGACTTGAAGAACTCGCGGAAGAA	148
QY	121	TGCTTTCTTTGAAATTTTGGAAACAGAAAGATCATGCTCGATCAAAAGTGGAAACGCTG	180
DB	149	TGCTTTCTTTGAAATTTTGGAAACAGAAAGATCATGCTCGATCAAAAGTGGAAACGCTG	208
QY	181	AAGCCATATCATGCTCATAAAGAGAAATGATAAAATTAACAGGGTAACCATTCCTCAG	240
DB	209	AAGCCATATCATGCTCATAAAGAGAAATGATAAAATTAACAGGGTAACCATTCCTCAG	268
QY	241	CAAGCGGTAGATGCTGTGCAAGAGATTCTCTCAGAGAGCCAAAGGAAAGACAGCTCA	300

```
|||||
Db 269 CAAGCGGTAGTCTGTCGAAGAGTTCCTCAGGAGAGCCAAAGGAAACACAGACGCTCA 328
QY 301 TCCCAAAATCTCTTGATGACAAAGATGACAGTAATTCAGTGAAGAGAGTAGGCGCA 360
Db 329 TCCCAAAATCTCTTGATGACAAAGATGACAGTAATTCAGTGAAGAGAGTAGGCGCA 388
QY 361 AACTCAGGTGATGAGAAGCGCAAACTTAGCCTGTCTGGAAGGAAGGTGAAGAAACATG 420
Db 389 AACTCAGGTGATGAGAAGCGCAAACTTAGCCTGTCTGGAAGGAAGGTGAAGAAACATG 448
QY 421 GGAGAAGGAAGAGAGGTGTCTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTG 480
Db 449 GGAGAAGGAAGAGAGGTGTCTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTG 508
QY 481 AAAAGAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTCCGCCCCCAAGAGATGAGAAGGAT 540
Db 509 AAAAGAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTCCGCCCCCAAGAGATGAGAAGGAT 568
QY 541 CTCACCATCCCGAGTCTAGTACCGTGAAGAGGGATGATGGCCGAGCCGATGCGCGCTTT 600
Db 569 CTCACCATCCCGAGTCTAGTACCGTGAAGAGGGATGATGGCCGAGCCGATGCGCGCTTT 628
QY 601 AAATGCGACCAACCGCAAGCGAGCTGTAAAGATGAGATCCTCATTTCCATCATTTTC 660
Db 629 AAATGCGACCAACCGCAAGCGAGCTGTAAAGATGAGATCCTCATTTCCATCATTTTC 688
QY 661 CTGCTAAGCCAAACAGAGAGAGCGAGCTGTCTGTACCAGGCAATCACGAAGAGTTGAAA 720
Db 689 CTGCTAAGCCAAACAGAGAGAGCGAGCTGTCTGTACCAGGCAATCACGAAGAGTTGAAA 748
QY 721 ATATGTGAAGA 731
Db 749 ATATGTGAAGA 759

B0720854 870 bp mRNA linear EST 16-JUL-2002
AGENCOURT 8241262 Lupski_sympathetic_trunk Homo sapiens cDNA clone
IMAGE:6187142 5', mRNA sequence.
B0720854
B0720854.1 GI:21859751
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Dr. James R. Lupski
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13580 row: k column: 15
High quality sequence stop: 668.
Location/Qualifiers
1. .870
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6187142"
/sex="male"
/tissue_type="sympathetic trunk"
/dev_stage="adult, 16 yr"
/lab_host="DH10B"

FEATURES
source
```

```
/clone_lib="Lupski_sympathetic_trunk"
/Note="vector: pCMV-SPORT6 (Life Technologies); Site_1:
NotI; Site_2: SalI; cDNA made by oligo-dT priming.
Directionally cloned using the following adaptors:
5'-TCGACCCACGCTCG-3' and
5'-GACTAGTCTAGATCGGAGGCGCCCT(15)-3'. Size selected >
1 kb for average insert length 1.9 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine); available through Life
Technologies."
BASE COUNT 270 a 193 c 245 g 161 t 1 others
ORIGIN
Query Match 97.3%; Score 727; DB 13; Length 870;
Best Local Similarity 100.0%; Pred. No. 1.1e-180;
Matches 727; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 CGGCTGTGAGTCTCGGCTCGGCGACTTGGTGTGGGGGAAACTCGGCCGATATCTCTCTT 64
Db 1 CGGCTGTGAGTCTCGGCTCGGCGACTTGGTGTGGGGGAAACTCGGCCGATATCTCTCTT 60
QY 65 GGCCAGGAAGATTTGTAATCCACCAGAGGACTTCAAGAAACCTCCGGAAGAAATGCT 124
Db 61 GGCCAGGAAGATTTGTAATCCACCAGAGGACTTCAAGAAACCTCCGGAAGAAATGCT 120
QY 125 TCTTTGTCAATTTTTTGGACAGAGATCATGCTCGCTGGATCAAAAGTGAACAGCTGAAGC 184
Db 121 TCTTTGTCAATTTTTTGGACAGAGATCATGCTCGCTGGATCAAAAGTGAACAGCTGAAGC 180
QY 185 CATATCATGCTCATAAAGAGAAATGATAAAATTAACAAGGGTAAACGATTCCAGCAAG 244
Db 181 CATATCATGCTCATAAAGAGAAATGATAAAATTAACAAGGGTAAACGATTCCAGCAAG 240
QY 245 CGGTAGATGCTGTGGAAGAGTTCCTCAGGAGAGCCAAAGGGAAAGACCAGAGCTCATCCC 304
Db 241 CGGTAGATGCTGTGGAAGAGTTCCTCAGGAGAGCCAAAGGGAAAGACCAGAGCTCATCCC 300
QY 305 ACAATTTCTTGATGACAAAGATCCAGCTAATTCAGTGAAGAGAGAGTAGGCCAAATCT 364
Db 301 ACAATTTCTTGATGACAAAGATCCAGCTAATTCAGTGAAGAGAGAGTAGGCCAAATCT 360
QY 365 CAGGTGATGAGAAGCGCAAACTTAGCCTGTCTGAAGGGAAGGTGAAGAAACATGGGAG 424
Db 361 CAGGTGATGAGAAGCGCAAACTTAGCCTGTCTGAAGGGAAGGTGAAGAAACATGGGAG 420
QY 425 AAGGAAGAGAGGGGTGTCTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTGAAA 484
Db 421 AAGGAAGAGAGGGGTGTCTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTGAAA 480
QY 485 GAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTCCGCCCCCAAGAGATGAGAAGGATCTCA 544
Db 481 GAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTCCGCCCCCAAGAGATGAGAAGGATCTCA 540
QY 545 CCATCCCGAGTCTAGTACCGTGAAGGGGATGATCGCCGAGCCGATGGCCGCTTTAAAT 604
Db 541 CCATCCCGAGTCTAGTACCGTGAAGGGGATGATGGCCGAGCCGATGGCCGCTTTAAAT 600
QY 605 GGCAGCAACCGCAAGAGGCTGTAAAGATGAGATCCTCATTTCCATCATTTCTGTC 664
Db 601 GGCAGCAACCGCAAGAGGCTGTAAAGATGAGATCCTCATTTCCATCATTTCTGTC 660
QY 665 TAAGCCAAACAGAGAGCCAGCTGTCTGTTACCAGCAATCACGAAGAGTTGAAAATAT 724
Db 661 TAAGCCAAACAGAGAGCCAGCTGTCTGTTACCAGCAATCACGAAGAGTTGAAAATAT 720
QY 725 GTGAAGA 731
Db 721 GTGAAGA 727

RESULT 4
BM808810
```

LOCUS BM808810 1039 bp mRNA linear EST 05-MAR-2002
 DEFINITION AGENCOURT 6617726 NIH_MGC_124 Homo sapiens cDNA clone IMAGE:5734324
 5', mRNA sequence.

ACCESSION BM808810
 VERSION BM808810.1 GI:19125633
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE 1. (bases 1 to 1039)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NIH-MGC http://mgi.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Invitrogen
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
 Plate: LLAM12739 row: h column: 05
 High quality sequence stop: 715.

FEATURES
 source
 1..1039
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5734324"
 /tissue_type="hippocampus"
 /lab_host="DH10B"
 /clone_lib="NIH_MGC_124"
 /notes="Organ: Brain; Vector: pCMV-SPORT6; Site 1: EcoRV (destroyed); Site 2: NotI; RNA source male hippocampus, age 27. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 0.9-4 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 012."

BASE COUNT 307 a 247 c 288 g 197 t

ORIGIN
 Query Match 97.2%; Score 726.4; DB 12; Length 1039;
 Best Local Similarity 99.9%; Pred. No. 1.6e-180;
 Matches 727; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 GCGGCTGTGAGTCTGCGGCTCGCGGACTTGTGTGGGGGAACTCGGCCGATATCCTCCT 63
 DB 9 GCGGCTGTGAGTCTGCGGCTCGCGGACTTGTGTGGGGGAACTCGGCCGATATCCTCCT 68
 QY 64 TGCCAGGAAAGATTGTTAATCCACCAAGGACTTGAAGAACTCGCGGAAAGAAATGC 123
 DB 69 TGCCAGGAAAGATTGTTAATCCACCAAGGACTTGAAGAACTCGCGGAAAGAAATGC 128
 QY 124 TTCTTTGTGAATTTTGTGAACAGAGATCATGCTCGGATCAAGTGGAAACAGCTGAAG 183
 DB 129 TTCTTTGTGAATTTTGTGAACAGAGATCATGCTCGGATCAAGTGGAAACAGCTGAAG 188
 QY 184 CCATATCATGCTCATAAAGAGGAAATGATAAAATTAACAGGTTAAACGATTCAGCAA 243
 DB 189 CCATATCATGCTCATAAAGAGGAAATGATAAAATTAACAGGTTAAACGATTCAGCAA 248
 QY 244 GCGGTAGATGCTGTGGAAGTTCCTCAGGAGAGCCAAAGGGAAGACCAAGCGTCATCC 303
 DB 249 GCGGTAGATGCTGTGGAAGTTCCTCAGGAGAGCCAAAGGGAAGACCAAGCGTCATCC 308
 QY 304 CACAATTCCTGATGACAGAAATCGACGTAATTCAGTGAGGAGAGAGTAGGCCAAAC 363
 DB 309 CACAATTCCTGATGACAGAAATCGACGTAATTCAGTGAGGAGAGAGTAGGCCAAAC 368

QY 364 TCAGGTGATGAGAGCGCAAACTTAGCTCTGTCTGAAGGGAAGGTGAAGAGAAATGGGA 423
 DB 369 TCAGGTGATGAGAGCGCAAACTTAGCTCTGTCTGAAGGGAAGGTGAAGAGAAATGGGA 428
 QY 424 GAAGGAAAGAGAGGGGTGCTTTCAGGCTCTTTCAGAGAGAGGCTCCAAATCCCTCTGAAA 483
 DB 429 GAAGGAAAGAGAGGGGTGCTTTCAGGCTCTTTCAGAGAGAGGCTCCAAATCCCTCTGAAA 488
 QY 484 AGAGGCCCAAGAGCAAAAGTCCCGGAAGCGGGTCCGCCCAAGGATGAGAAGGATCTC 543
 DB 489 AGAGGCCCAAGAGCAAAAGTCCCGGAAGCGGGTCCGCCCAAGGATGAGAAGGATCTC 548
 QY 544 ACCATCCCGAGTCTAGTACCGTGAAGGGGATGATGCCGACCGATGCGCGGTTTAAA 603
 DB 549 ACCATCCCGAGTCTAGTACCGTGAAGGGGATGATGCCGACCGATGCGCGGTTTAAA 608
 QY 604 TGCAGCCCAACCGCAAGCGAGCTGTTAAAGATGCGAGATCTCATTTCCATCATTTTCCTG 663
 DB 609 TGCAGCCCAACCGCAAGCGAGCTGTTAAAGATGCGAGATCTCATTTCCATCATTTTCCTG 668
 QY 664 CTAAGCCCAACAGAGAGCGAGCTGCTGTACCGCAATCACGAAGAGTTGAAATA 723
 DB 669 CTAAGCCCAACAGAGAGCGAGCTGCTGTACCGCAATCACGAAGAGTTGAAATA 728
 QY 724 TGTGAAGA 731
 DB 729 TGTGAAGA 736

RESULT 5
 LOCUS BUI58924
 DEFINITION AGENCOURT 7844927 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6044084
 5', mRNA sequence.

ACCESSION BUI58924
 VERSION BUI58924.1 GI:22672834
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE 1. (bases 1 to 884)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NIH-MGC http://mgi.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
 Plate: LLAM13286 row: b column: 21
 High quality sequence stop: 776.

FEATURES
 source
 1..884
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6044084"
 /tissue_type="embryonal carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_92"
 /note="Organ: testis; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 2.5 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC library." 274 a 195 c 244 g 169 t 2 others

BASE COUNT 274 a 195 c 244 g 169 t 2 others
 ORIGIN

```
Query Match 97.2%; Score 726; DB 13; Length 884;
Best Local Similarity 100.0%; Pred. No. 2e-180;
Matches 726; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GCGTGTGAGTCTCGCGCTCGCGACTTGTGTGGGGAACCTCGCGGATATCCCTCTTG 65
DB 1 GCGTGTGAGTCTCGCGCTCGCGACTTGTGTGGGGAACCTCGCGGATATCCCTCTTG 60
QY 66 GCGAGAAAGATTGTTAATCCCAAGGACTTTGAAGAAACCTCGCGGAAAGAAATGCTT 125
DB 61 GCGAGAAAGATTGTTAATCCCAAGGACTTTGAAGAAACCTCGCGGAAAGAAATGCTT 120
QY 126 CTTTGTGAAATTTTGGAAACAGAGATCATGCTGGATCAAGTGAAGTGAACAGCTGAAGCC 185
DB 121 CTTTGTGAAATTTTGGAAACAGAGATCATGCTGGATCAAGTGAAGTGAACAGCTGAAGCC 180
QY 186 ATATCATGCTCATAAAGAGGAATGATAAAATTACAAAGGTAAACGATTCCAGCAAGC 245
DB 181 ATATCATGCTCATAAAGAGGAATGATAAAATTACAAAGGTAAACGATTCCAGCAAGC 240
QY 246 GGTAGATGCTGTGCAAGAGTTCTCTAGGAGAGCCAAAGGAAAGACGACGTCATCCCA 305
DB 241 GGTAGATGCTGTGCAAGAGTTCTCTAGGAGAGCCAAAGGAAAGACGACGTCATCCCA 300
QY 306 CAATTCCTTGATGACAAAGATCGAGTAAATCCAGTGAAGAGAGAGTGAAGCAAACTC 365
DB 301 CAATTCCTTGATGACAAAGATCGAGTAAATCCAGTGAAGAGAGAGTGAAGCAAACTC 360
QY 366 AGTGATGAGAGAGCGCAAACTTAGCTGTCTGAAGGAGGTGAAGAAACATGGGAGA 425
DB 361 AGTGATGAGAGAGCGCAAACTTAGCTGTCTGAAGGAGGTGAAGAAACATGGGAGA 420
QY 426 AGGAAAGAGAGAGGTGCTTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTGAAAAG 485
DB 421 AGGAAAGAGAGAGGTGCTTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTGAAAAG 480
QY 486 AGCCCAAGAGCAAAAGTCCCGGAGCGGGTGGCCCCCAAGAGATGAAGAGATCTCAC 545
DB 481 AGCCCAAGAGCAAAAGTCCCGGAGCGGGTGGCCCCCAAGAGATGAAGAGATCTCAC 540
QY 546 CATCCCGGAGTCTAGTACCGTGAAGGGGATGATGGCCGAGCCGATGGCCGCTTTAATG 605
DB 541 CATCCCGGAGTCTAGTACCGTGAAGGGGATGATGGCCGAGCCGATGGCCGCTTTAATG 600
QY 606 GCAGCCAAACCGCAAGCGAGCTTTAAAGATGAGATCTCTCATTTCCATCTTCTGCT 665
DB 601 GCAGCCAAACCGCAAGCGAGCTTTAAAGATGAGATCTCTCATTTCCATCTTCTGCT 660
QY 666 AAGCCAAACAGAGAGCCAGCTGTCTGTTACAGGCAATCAGAGAGAGTGAAGAAATG 725
DB 661 AAGCCAAACAGAGAGCCAGCTGTCTGTTACAGGCAATCAGAGAGAGTGAAGAAATG 720
QY 726 TGAAGA 731
DB 721 TGAAGA 726

RESULT 6
AL536040
LOCUS
DEFINITION
AL536040 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
CS0DF022Y005 5-PRIME, mRNA sequence.
ACCESSION
AL536040
VERSION
AL536040.2 GI:31260910
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1003)
Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
```

COMMENT

On Feb 13, 2001 this sequence version replaced gi:12799533.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segre@genoscope.cns.fr, Web: www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 7982.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DF022AH030P1&cluster=7982.r. Contact :
Peng Liang Email: fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DF022AH030P1.

FEATURES

source

Location/Qualifiers
1..1003
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DF022Y005"
/tissue_type="FETAL BRAIN"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL BRAIN"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."

BASE COUNT 308 a 221 c 273 g 196 t

ORIGIN

Query Match 96.1%; Score 718; DB 9; Length 1003;
Best Local Similarity 99.2%; Pred. No. 2.7e-178;
Matches 726; Conservative 5; Mismatches 0; Indels 1; Gaps 1;

QY 1 ATGGCGGCTGTGAGTCTCGCGCTCGCGACTTGTGTGGG-GGAACTCGCGGATATCC 59
DB 64 ATGGCGGCTGTGAGTCTCGCGCTCGCGACTTGTGTGGG-GGAACTCGCGGATATCC 123
QY 60 TCCTTGGCCAGAAAGATTGTTAATCCCAAGAGACTTGAAGAAACCTCGCGGAAAGAA 119
DB 124 TCCTTGGCCAGAAARATGTTAATCCCAAGAGACTTGAAGAAACCTCGCGGAAAGAA 183
QY 120 ATGCTTCTTTGAAATTTTGGAAACAGATCATGCTCGATCAAAAGTGAACAGCT 179
DB 184 ATGCTTCTTTGAAATTTTGGAAACAGATCATGCTCGATCAAAAGTGAACAGCT 243
QY 180 GAAGCCATATCATGCTCATAAAGAGGAATGATAAAATTAAACAGGTAAACGATTCCA 239
DB 244 GAAGCCATATCATGCTCATAAARAGGAATGATAAAATTAAACAGGTAAACGATTCCA 303
QY 240 GCAAGCGGTAGATGCTGTGCAAGAGTTCTCTAGGAGAGCCAAAGGAAAGACAGACGTC 299
DB 304 GCAAGCGGTAGATGCTGTGCAAGAGTTCTCTAGGAGAGCCAAAGGAAARACACAGTC 363
QY 300 ATCCCACAATTTCTTGATGACAAAGATCGAGTAAATTCAGTGAAGAGAGTGAAGTGGCC 359
DB 364 ATCCCACAATTTCTTGATGACAAAGATCGAGTAAATTCAGTGAAGAGAGTGAAGTGGCC 423
QY 360 AAACCTCAGTGTATGAGAGCGCAAACTTAGCTGTCTGAAGGAGAGGTGAAGAGACAT 419
DB 424 AAACCTCAGTGTATGAGAGCGCAAACTTAGCTGTCTGAAGGAGAGGTGAAGAGACAT 483
QY 420 GGGAGAGGAAAGAGAGGGTGTCTTTACGGCTCTTCAGAGAGAGGCTCCAAATCCCTCT 479
DB 484 GGGAGAGGAAARAGAGGGTGTCTTTACGGCTCTTCAGAGAGAGGCTCCAAATCCCTCT 543
QY 480 GAAAGAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTCCGCCCCCAAGAGATGAAGAGA 539
DB 544 GAAAGAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTCCGCCCCCAAGAGATGAAGAGA 603
QY 540 TCTCACCATCCCGAGTCTAGTACCGTGAAGGGGATGATGGCCGAGACCGATGGCCGCTT 599
DB 604 TCTCACCATCCCGAGTCTAGTACCGTGAAGGGGATGATGGCCGAGACCGATGGCCGCTT 663

```

600 TAAATGGCAGCCAAACCGCAGCGCTGTTAAAGATGCAGATCCTCATTTCCATCATTT 659
Db      |||||||
664 TAAATGGCAGCCAAACCGCAGCGCTGTTAAAGATGCAGATCCTCATTTCCATCATTT 723
Qy      |||||||
660 CTGCTTAAGCCAAACAGAGAGCCAGCTGCTGTACCGCAATCACGAAGAAGTTGAA 719
Db      |||||||
724 CTTGCTAAGCCAAACAGAGAGCCAGCTGCTGTACCGCAATCACGAAGAAGTTGAA 783
Qy      |||||||
720 AATATGTGAAGA 731
Db      |||||||
784 AATATGTGAAGA 795

RESULT 7
BM808969
LOCUS      6617730 NIH_MGC_124 Homo sapiens cDNA clone IMAGE:5734420
DEFINITION      5', mRNA sequence.
ACCESSION      BM808969
VERSION        1
KEYWORDS       EST.
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
REFERENCE      1 (bases 1 to 1005)
AUTHORS       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE         Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL       NIH-MGC http://mgc.nci.nih.gov/.
COMMENT       National Institutes of Health, Mammalian Gene Collection (MGC)
              Unpublished
              Contact: Robert Strausberg, Ph.D.
              Email: cgabbs-r@mail.nih.gov
              Tissue Procurement: Invitrogen
              CDNA Library Preparation: Life Technologies, Inc.
              CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Agencourt Bioscience Corporation
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLAM12739 Row: 1 column: 05
              High quality sequence stop: 639.
              Location/Qualifiers
                1..1005
                  /organism="Homo sapiens"
                  /mol_type="mRNA"
                  /db_xref="taxon:9606"
                  /clone="IMAGE:5734420"
                  /tissue_type="hippocampus"
                  /lab_host="DH10B"
                  /clone_lib="NIH MGC 124"
                  /notes="Organ: brain; Vector: pCMV-SPORT6; Site 1: EcoRV
                  (destroyed); Site 2: NotI; RNA source male hippocampus,
                  age 27. Library is oligo-dT primed and directionally
                  cloned (EcoRV site is destroyed upon cloning). Average
                  insert size 1.4 kb, insert size range 0.9-4 kb. Library is
                  normalized and enriched for full-length clones and was
                  constructed by C. Gruber (Invitrogen). Research Genetics
                  tracking code 012."
BASE COUNT      299 a 235 c 278 g 193 t
ORIGIN
Query Match      93.8%; Score 700.4; DB 12; Length 1005;
Best Local Similarity 98.9%; Pred. No. 1.2e-173;
Matches 716; Conservative 0; Mismatches 6; Indels 2; Gaps 1;

Qy 1 ATGCGCGCTGTAGCTCGCGCTCGCGACTTGTGTGGGGAAACTCGCGCATATCCT 60
Db  |||||||
20 ATGCGCGCTGTAGCTCGCGCTCGCGACTTGTGTGGGGAAACTCGCGCATATCCT 79
Qy 61 CCTTGGCCAGGAAGATTGTAATCCACCAAGGACTTGAAGAACTCGCGGAAGAA 120
Db  |||||||
80 CCTTGGCCAGGAAGATTGTAATCCACCAAGGACTTGAAGAACTCGCGGAAGAA 139

```

```

Qy 121 TGCCTCTTTGTGAAATTTTTTGGAAACAGAGATCATGCTCGATCAAGTGGAAACAGCTG 180
Db  |||||||
140 TGCCTCTTTGTGAAATTTTTTGGAAACAGAGATCATGCTCGATCAAGTGGAAACAGCTG 199
Qy 181 AGCCCATATCATGCTCATAAAGAGGAATGATAAAATTTAAACAGGTTAAACGATTTCCAG 240
Db  |||||||
200 AGCCCATATCATGCTCATAAAGAGGAATGATAAAATTTAAACAGGTTAAACGATTTCCAG 259
Qy 241 CAAAGCGGTAGATGCTGTCGAAGAGATTCTCTCAGGAGAGCCAAAGGGGAAAGACACGATCA 300
Db  |||||||
260 CAAAGCGGTAGATGCTGTCGAAGAGATTCTCTCAGGAGAGCCAAAGGGGAAAGACACGATCA 319
Qy 301 TCCCAACAATTTCTTCGATGACCAAGATCGACGTAATTCAGTGAGAGAGAGATAGGCCA 360
Db  |||||||
320 TCCCAACAATTTCTTCGATGACCAAGATCGACGTAATTCAGTGAGAGAGAGATAGGCCA 379
Qy 361 AACTCAGGTGATGAGAGAGCCAAACTTAGCCTGTCTGAAGGGAGAGTGAAGAGAACATG 420
Db  |||||||
380 AACTCAGGTGATGAGAGAGCCAAACTTAGCCTGTCTGAAGGGAGAGTGAAGAGAACATG 439
Qy 421 GGAGAGAGGAAAGAGAGGGGTGTCTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTG 480
Db  |||||||
440 GGAGAGAGGAAAGAGAGGGGTGTCTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTG 499
Qy 481 AAAAGAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTCCGCCCAAGAGATGAGAGGAT 540
Db  |||||||
500 AAAAGAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTCCGCCCAAGAGATGAGAGGAT 559
Qy 541 CTCACCATCCCGAGTCTAGTACCGTGAAGGGAGATGATGCCGGACCGATGGCCGCTTT 600
Db  |||||||
560 CTCACCATCCCGAGTCTAGTACCGTGAAGGGAGATGATGCCGGACCGATGGCCGCTTT 619
Qy 601 AAATGCGACGCAACCGCAAGCGAGCTGTTAAAGATGCAGATCCTCATTTCCATCATTTTC 660
Db  |||||||
620 AAATGCGACGCAACCGCAAGCGAGCTGTTAAAGATGCAGATCCTCATTTCCATCATTTTC 679
Qy 661 CTGCTAAGCCCAACAGAGAGAGCGAGCTGTGTACAGGC--AATCAAGAAAGATTGA 718
Db  |||||||
680 CTGCTAAGCCCAACAGAGAGAGCGAGCTGTGTACAGGC--AATCAAGAAAGATTGA 739
Qy 719 AAAT 722
Db  |||||
740 AAAT 743

RESULT 8
BM851885
LOCUS      603192793F1 NTH_MGC_95 Homo sapiens cDNA clone IMAGE:5263991 5',
DEFINITION      mRNA sequence.
ACCESSION      BI551885
VERSION        BI551885.1 GI:15439197
KEYWORDS       EST.
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
              1 (bases 1 to 755)
              NIH-MGC http://mgc.nci.nih.gov/.
              National Institutes of Health, Mammalian Gene Collection (MGC)
              Unpublished
              Contact: Robert Strausberg, Ph.D.
              Email: cgabbs-r@mail.nih.gov
              Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
              CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
              Toshiyuki and Piero Carninci (RIKEN)
              CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLAM11665 Row: j column: 24
              High quality sequence stop: 746.

```

FEATURES
source

Location/Qualifiers
1. .755
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5263991"
/tissue_type="hippocampus"
/lab_host="DH10B"
/clone_lib="NIH_MGC_95"
/note="Organ: brain; Vector: pBluescriptR (modified pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtcgag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.5 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT 236 a 156 c 218 g 145 t

ORIGIN

Query Match 92.7%; Score 692.8; DB 12; Length 755;
Best Local Similarity 99.3%; Pred. No. 1.1e-171;
Matches 727; Conservative 0; Mismatches 2; Indels 3; Gaps 3;

QY 1 ATGGCGGCTGTGAGTCTGGCGCTCGCGGCTTGGTGTGGGGAACTCGGCCGATATCCT 60
DB 25 ATGGCGGCTGTGAGTCTGGCGCTCGCGGCTTGGTGTGGGGAACTCGGCCGATATCCT 84
QY 61 CTTTGGCCAGGAAGATTGTTAATCAACCAAGGACTTGAAGAACTCGCGGAAGAAA 120
DB 85 CTTTGGCCAGGAAGATTGTTAATCAACCAAGGACTTGAAGAACTCGCGGAAGAAA 144
QY 121 TGGCTCTTTGTGAAATTTTGAACAGAGATCATGCTCGTGAATCAAGTGAACAGCTG 180
DB 145 TGGCTCTTTGTGAAA-TTTTGAACAGAGATCATGCTCGTGAATCAAGTGAACAGCTG 203
QY 181 AAGCCATATCATGCTCATAAAGAGGAATGATAAAATTAACAAAGGTTAAACGATTCAG 240
DB 204 AAGCCATATCATGCTCATAAAGAGGAATGATAAAATTAACAAAGGTTAAACGATTCAG 263
QY 241 CAAGCGGTAGATGCTGTGGAAGAGTTCTTCAGAGAGCCAAAGGGAAAGACAGAGTCA 300
DB 264 CAAGCGGTAGATGCTGTGGAAGAGTTCTTCAGAGAGCCAAAGGGAAAGACAGAGTCA 323
QY 301 TCCCAATCTTCTGATGACAGATCGACGTAATTCAGTCAGGAGAGAGTAGGCA 360
DB 324 TCCCAATCTTCTGATGACAGATCGACGTAATTCAGTCAGGAGAGAGTAGGCA 383
QY 361 AACTCAGGTGATGAGAGGCAAACTTAGCTCTCTGAAGGGAGGTGAAGAGAACATG 420
DB 384 AACTCAGGTGATGAGAGGCAAACTTAGCTCTCTGAAGGGAGGTGAAGAGAACATG 443
QY 421 GGAGAGGAAGAGAGGTTGTTCTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTG 480
DB 444 GGAGAGGAAGAGAGGTTGTTCTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTG 503
QY 481 AAAAGAGCCCAAGAGCAAGTCCCGAGACGGGGTCGGCCCCCAAGAGATGAGAGAT 540
DB 504 AAAAGAGCCCAAGAGCAAGTCCCGAGACGGGGTCGGCCCCCAAGAGATGAGAGAT 563
QY 541 CTCACCATCCCGAGTCTAGTACCGTGAAGGGATGATGCGCGGACCGATGCGCGGTTT 600
DB 564 CTCACCATCCCGAGTCTAGTACCGTGAAGGGATGATGCGCGGACCGATGCGCGGTTT 623
QY 601 AAATGGCAGCAACCGCAAGCGAGCTGTTTAAAGATGACAGATCCTCATTTCCATCA-TTT 659
DB 624 AAATGGCAGCAACCGCAAGCGAGCTGTTTAAAGATGACAGATCCTCATTTCCATCA-TTT 683
QY 660 CTTGCTTAAGCC-AAACAGAGAACCCAGCTGTCTGTGTACAGGCAATCAAGAGAGTTGA 718
DB 684 CTTGCTTAAGCCAAACAGAGAACCCAGCTGTCTGTGTACAGGCAATCAAGAGAGTTGA 743

QY 719 AAATATGTGAAG 730
DB 744 AAATATGTGAAG 755

RESULT 9

BI600148

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

778 bp mRNA linear EST 07-SEP-2001
603246545F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5288915 5',
mRNA sequence.
BI600148
BI600148.1 GI:15493087
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 778)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D. (NHGRI), Shiraki
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11730 row: i column: 12
High quality sequence stop: 773.

FEATURES

source

1. .778
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5288915"
/tissue_type="hypothalamus"
/lab_host="DH10B"
/clone_lib="NIH_MGC_96"
/note="Organ: brain; Vector: pBluescriptR (modified pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtcgag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.3 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT 237 a 162 c 227 g 152 t

ORIGIN

Query Match 90.1%; Score 673; DB 12; Length 778;
Best Local Similarity 98.2%; Pred. No. 1.8e-166;
Matches 712; Conservative 0; Mismatches 10; Indels 3; Gaps 3;

QY 1 ATGGCGGCTGTGAGTCTGGCGCTCGCGGCTTGGTGTGGGGAACTCGGCCGATATCCT 60
DB 26 ATGGCGGCTGTGAGTCTGGCGCTCGCGGCTTGGTGTGGGGAACTCGGCCGATATCCT 85
QY 61 CTTTGGCCAGGAAGATTGTTAATCAACCAAGGACTTGAAGAACTCGCGGAAGAAA 120
DB 86 CTTTGGCCAGGAAGATTGTTAATCAACCAAGGACTTGAAGAACTCGCGGAAGAAA 145
QY 121 TGGCTCTTTGTGAAATTTTGGAAACAGAGATCATGCTCGGATCAAGTGAACAGCTG 180
DB 146 TGGCTCTTTGTGAAATTTTGGAAACAGAGATCATGCTCGGATCAAGTGAACAGCTG 205
QY 181 AAGCCATATCATGCTCATAAAGAGGAATGATAAAATTAACAAAGGTTAAACGATTCAG 240
DB 206 AAGCCATATCATGCTCATAAAGAGGAATGATAAAATTAACAAAGGTTAAACGATTCAG 265

```
QY 241 CAACGGGTAGATGCTGTGCGAGAGTTCTCTCAGGAGAGCCAAAGGAAAGACACGACGTCA 300
    |||
Db 266 CAACGGGTAGATGCTGTGCGAGAGTTCTCTCAGGAGAGCCAAAGGAAAGACACGACGTCA 325
    |||
QY 301 TCCACAAATTTCTTGATGACAAAGATCGAGTAAATTCAGTGAAGGAGAGATGAGGCCA 360
    |||
Db 326 TCCACAAATTTCTTGATGACAAAGATCGAGTAAATTCAGTGAAGGAGAGATGAGGCCA 385
    |||
QY 361 AACTCAGGTGATGAGAGCGCAAACTTAGCTGTCTGAAGGAGAGGTGAAGAAACATG 420
    |||
Db 386 AACTCAGGTGATGAGAGCGCAAACTTAGCTGTCTGAAGGAGAGGTGAAGAAACATG 445
    |||
QY 421 GGAGAGGAAAGAGAGGGTGTCTTTCAGGCTCTTTCAGAGAGAGGCTCCAAATCCCTCTG 480
    |||
Db 446 GGAGAGGAAAGAGAGGGTGTCTTTCAGGCTCTTTCAGAGAGAGGCTCCAAATCCCTCTG 505
    |||
QY 481 AARAGAGCCCAAGAGCAAAAGTCCCGGAGAGGGGTGCGGCCCAAGAGATGAGAAGAT 540
    |||
Db 506 AARAGAGCCCAAGAGCAAAAGTCCCGGAGAGGGGTGCGGCCCAAGAGATGAGAAGAT 564
    |||
QY 541 CTACCAATCCGGAGTCTA-GTACCGTGAAGGGGATGATGCGCGACCGATGCGCGGTT 599
    |||
Db 565 CTACCAATCCGGAGTCTAGTACCGTGAAGGGGATGATGCGCGACCGATGCGCGGTT 624
    |||
QY 600 TAAATGCGAGCCCAACCGCAAGCGAGCTGTGTTAAAGATGCGAGTCTCTCAATTTCCATCATTT 659
    |||
Db 625 TAAATGCGAGCCCAACCGCAAGCGAGCTGTGTTAAAGATGCGAGTCTCTCAATTTCCATCATTT 684
    |||
QY 660 CTGCTGAAGCAAAACAGAGAGCCAGCTGTCTGTTACGAGCAATACAGAAAGTTGAA 719
    |||
Db 685 CTGCTGAAGCAAAACAGAGAGCCAGCTGTCTGTTACGAGC-ATCAGGAAGATTGAA 743
    |||
QY 720 AATAT 724
    ||||
Db 744 AATAT 748
    ||||
```

```
RESULT 10
LOCUS BI459800
DEFINITION BI459800 816 bp mRNA linear EST 21-AUG-2001
mRNA sequence.
ACCESSION BI459800
VERSION BI459800.1 GI:15250456
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NTH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LLAM11672 row: h column: 13
High quality sequence stop: 719.
Location/Qualifiers
1. .816
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:526620"
/lab_host="DH10B"
```

FEATURES

source

RESULT 11

BM906992

LOCUS

1138 bp mRNA linear EST 12-MAR-2002

```
BASE COUNT 254 a 181 c 230 g 151 t
ORIGIN
Query Match 89.5%; Score 668.8; DB 12; Length 816;
Best Local Similarity 97.9%; Pred. No. 2.4e-165;
Matches 709; Conservative 0; Mismatches 12; Indels 3; Gaps 3;
QY 1 ATGGCGGCTGTGAGTCTGCGGCTCGCGGCTGCTGTTGGGGGAACTCGGCCCATATCTCT 60
Db 9 ATGGCGGCTGTGAGTCTGCGGCTCGCGGCTGCTGTTGGGGGAACTCGGCCCATATCTCT 68
    |||
QY 61 CCTTGGCCAGGAAAGATTTGTTAATCCACAAAGGACTTGAAGAAACCTCGCGGAAAGAA 120
    |||
Db 69 CCTTGGCCAGGAAAGATTTGTTAATCCACAAAGGACTTGAAGAAACCTCGCGGAAAGAA 127
    |||
QY 121 TGCCTCTTTGTGAAATTTTTTGGAAACAGAGATCATGCTCGGATCAAAAGTGGAAACAGCTG 180
    |||
Db 128 TGCCTCTTTGTGAAATTTTTTGGAAACAGAGATCATGCTCGGATCAAAAGTGGAAACAGCTG 187
    |||
QY 181 AAGCCATATCATGCTCATAAAGAGGAAATGATAAAATTAACNAGGGTAAACGATTCAG 240
    |||
Db 188 AAGCCATATCATGCTCATAAAGAGGAAATGATAAAATTAACNAGGGTAAACGATTCAG 247
    |||
QY 241 CAAGCGGTAGATGCTGTGCGAGAGTTCTCTCAGGAGAGCCAAAGGAAAGACAGAGCTCA 300
    |||
Db 248 CAAGCGGTAGATGCTGTGCGAGAGTTCTCTCAGGAGAGCCAAAGGAAAGACAGAGCTCA 307
    |||
QY 301 TCCACAAATTTCTTCTGATGACAAAGATTCACGCTAATTTCCAGTGAGGAGAGAGTAGGCCA 360
    |||
Db 308 TCCACAAATTTCTTCTGATGACAAAGATTCACGCTAATTTCCAGTGAGGAGAGAGTAGGCCA 367
    |||
QY 361 AACTCAGGTGATGAGAGCGCAAACTTAGCTGTCTGAAGGAGAGGTGAAGAGAACATG 420
    |||
Db 368 AACTCAGGTGATGAGAGCGCAAACTTAGCTGTCTGAAGGAGAGGTGAAGAGAACATG 427
    |||
QY 421 GGAGAGGAAAGAGAGGGTGTCTTTCAGAGAGAGGCTCCAAATCCCTCTG 480
    |||
Db 428 GGAGAGGAAAGAGAGGGTGTCTTTCAGAGAGAGGCTCCAAATCCCTCTG 487
    |||
QY 481 AAAAGAGCCCAAGAGCAAAAGTCCCGGAAAGCGGGTCCGCCCAAGGATGAGAAGGAT 540
    |||
Db 488 AAAAGAGCCCAAGAGCAAAAGTCCCGGAAAGCGGGTCCGCCCAAGGATGAGAAGGAT 546
    |||
QY 541 CTCACCATCCCGAGTCTAGTACCGTGAAGGGGATGATGCGCGGACCGATGCGCGGTTT 600
    |||
Db 547 CTCACCGTCCCGAGTCTAGTACCGTGAAGGGGATGATGCGCGGACCGATGCGCGGTTT 606
    |||
QY 601 AAATGCGAGCAACCGCAAGCGAGCTGTTAAAGATGCGAGTCTCATTTCCATCATTTTC 660
    |||
Db 607 AAATGCGAG-CAACGCAAGCGAGCTGTTAAAGATGCGAGTCTCATTTCCATCATTTTC 665
    |||
QY 661 CTGCTAAGCAAAACAGAGAGGAGCGAGCTGTCTGTTCAGGCAATACGAGAGAGTTGAA 720
    |||
Db 666 CTGCTAAGCAAAACAGAGAGGAGCGAGCTGTCTGTTCAGGCAATACGAGAGAGTTGACC 725
    |||
QY 721 AATAT 724
    |||
Db 726 AATT 729
    |||
```

BM906992

1138 bp

mRNA

linear

EST 12-MAR-2002


```

DEFINITION AGENCOURT_6622587 NIH_MGC_124 Homo sapiens cDNA clone IMAGE:5735036
5', mRNA sequence.
ACCESSION BM906992
VERSION BM906992.1 GI:19357371
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1138)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Invitrogen
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12741 row: e column: 21
High quality sequence stop: 444.

FEATURES
source
    1..1138
        Location/Qualifiers
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:5735036"
            /tissue_type="hippocampus"
            /lab_host="DH10B"
            /clone_lib="NIH MGC 124"
            /note="Organ: brain; Vector: pCMV-SPORT6; Site: 1: EcoRV
            (destroyed); Site: 2: NotI; RNA source male hippocampus,
            age 27. Library is oligo-dT primed and directionally
            cloned (EcoRV site is destroyed upon cloning). Average
            insert size 1.4 kb, insert size range 0.9-4 kb. Library is
            normalized and enriched for full-length clones and was
            constructed by C. Gruber (Invitrogen). Research Genetics
            tracking code 012."
BASE COUNT 309 a 323 c 289 g 217 t
ORIGIN
    Query Match 86.6%; Score 647; DB 12; Length 1138;
    Best Local Similarity 97.6%; Pred. No. 1.4e-159;
    Matches 678; Conservative 0; Mismatches 15; Indels 2; Gaps 2;

QY 1 ATGGCGGCTGTGAGTCTGGCGCTCGCGGACTTGGTGTGGGGAAACTCGCGGATATCTCT 60
DB 20 ATGGCGGCTGTGAGTCTGGCGCTCGCGGACTTGGTGTGGGGAAACTCGCGGATATCTCT 79
QY 61 CTTTGCCAGGAAGATTGTTATCCACCAAGGACTTGAAGAACTTCGCGGAAGAAA 120
DB 80 CTTTGCCAGGAAGATTGTTATCCACCAAGGACTTGAAGAACTTCGCGGAAGAAA 139
QY 121 TCGCTCTTGTGAAATTTTTCGACAGAGATCATGCTCGATCAAGTGGACAGCTG 180
DB 140 TCGCTCTTGTGAAATTTTTCGACAGAGATCATGCTCGATCAAGTGGACAGCTG 199
QY 181 AAGCCATATCATCTCTATGAAGAGGAAATGATATAAAATTAACAAAGGTTAAACGATTCCAG 240
DB 200 AAGCCATATCATCTCTATGAAGAGGAAATGATATAAAATTAACAAAGGTTAAACGATTCCAG 259
QY 241 CAAGCGGTAGATCTCTGCAAGAGTTCTTCAGAGAGCCAAAGGGAAGACCAAGCTCA 300
DB 260 CAAGCGGTAGATCTCTGCAAGAGTTCTTCAGAGAGCCAAAGGGAAGACCAAGCTCA 319
QY 301 TCCCACAATTTCTGATGACAGATTCGCGTAATTCAGTCAGGAGAGAGTAGGCA 360
DB 320 TCCCACAATTTCTGATGACAGATTCGCGTAATTCAGTCAGGAGAGAGTAGGCA 379
QY 361 AACTCAGGTGATGAGAAGCGCAAACTTAGCTCTGTGAAGGGAAGGTGAAGAAGACATG 420

```

```

DB 380 AACTCAGGTGATGAGAAGCGCAAACTTAGCTCTGTGAAGGGAAGGTGAAGAAGACATG 439
QY 421 GGAGAAGCAAGAAAGAGAGGTTGTCTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTG 480
DB 440 GGAGAAGCAAGAAAGAGAGGTTGTCTTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTG 499
QY 481 AAAAGAG-CCCAAGAGCAAAAGTCCCGAAGCGGGTCCGCCCCCAAAAGGATGAGAAGGA 539
DB 500 AAAAGAGCCCCAAGAGACAGTCCCGAAGCGGGCCGCCCCCAAAAGGATGAGAAGGA 559
QY 540 TCTACCATCCCGAGTCTAGTACCGTGAAGGGATGATGCCGACCGATGCCCGCT 599
DB 560 TCTACCATCCCGAGTCTAGTACCGTGAAGGGATGATGCCGACCGATGCCCGCT 619
QY 600 TAAATGGCAGCAACCGCAAGCGAGCTGTAAAG-ATGCAGATCTCATTTCCATCAT 658
DB 620 TTAATGGCAGCAACCGCAAGCGAGCTGTACAAATGGGATCTCATTTCCATCAT 679
QY 659 TCCTGCTAAGCCAAACAGAGAGCCAGCTGTCTGT 693
DB 680 TCCTGCTAACCACCAACCGAGAGGAATAATGGCTCT 714

RESULT 12
LOCUS BUS20128
DEFINITION AGENCOURT_10157423 NIH_MGC_134 Mus musculus linear EST 12-SEP-2002
IMAGE:6518047 5', mRNA sequence.
ACCESSION BUS20128
VERSION BUS20128.1 GI:22827654
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 993)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Dr. David Rowe
cDNA Library Preparation: Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM14099 row: c column: 08
High quality sequence stop: 675.

FEATURES
source
    1..993
        Location/Qualifiers
            /organism="Mus musculus"
            /mol_type="mRNA"
            /db_xref="taxon:10090"
            /clone="IMAGE:6518047"
            /tissue_type="undifferentiated limb"
            /lab_host="DH10B (phage-resistant)"
            /clone_lib="NIH MGC_134"
            /note="Vector: pCMV-SPORT6.1.ccd; Site: 1: EcoRV; Site: 2:
            NotI; Cloned unidirectionally. Primer: Oligo dT. Average
            insert size 1-7 kb. Constructed by ResGen, Invitrogen
            Corp. Note: this is a NIH_MGC Library."
BASE COUNT 293 a 222 c 275 g 194 t
ORIGIN
    Query Match 84.9%; Score 634.2; DB 13; Length 993;
    Best Local Similarity 92.9%; Pred. No. 3.3e-156;
    Matches 676; Conservative 0; Mismatches 49; Indels 3; Gaps 1;

QY 4 GCGGCTGTGAGTCTCGGCTCGCGGACTTGGTGTGGGGAAACTCGCGGATATCTCTCT 63

```

Db	1	CGGGCTGTAGTCTGCGGCTCGGCGACTTGGTGTGGGGAACTGGGCGGATCTCTCC	60
Qy	64	TGGCCAGGAAAGATTGTTAATCCACCAAGAGCTTTGAAGAAACCTCGCGGAAGAAATGC	123
Db	61	TGGCCAGGAAAGATTGTTAATCCACCAAGAGCTTTGAAGAAACCTCGCGGAAGAAATGC	120
Qy	124	TTCTTTGTGAAATTTTGTGAAACAGAAAGATCATGCTCGGATCAAAGTGAACAGCTGAAG	183
Db	121	TTCTTTGTGAAATTTTGTGAAACAGAAAGATCATGCTCGGATCAAAGTGAACAGCTGAAG	180
Qy	184	CCATATCATCTCATAAAGAGGAATGATAAATAATTAACAGGGTAACCAATTCAGCAAA	243
Db	181	CTTTACCATCTCATCAAGAGGAGATGATAAAGATTAAACAGGGTAACCAATTCAGCAAA	240
Qy	244	GCGGTAGATCTGTCGAAAGAGTTCCTCAGAGAGCCAAAGGGAAGACAGAGCTCATCC	303
Db	241	GCTGTGATCTGTTGAAGAGTTCCTCAGAGAGCCAAAGGGAAGACAGAGCTCATCC	300
Qy	304	CACAATTTCTTGATGACAAAGATTCGACGTAATTCAGTGTGAGGAGAAAGTAGGCCAAAC	363
Db	301	CACATTTCTGCTGATGACAAAGATTCGCGGTAATTCAGTGTGAGGAGAAAGTAGGCCAAAC	360
Qy	364	TCAGGTGATGAGAGCCAAACTAGCTCTCTGAAGGGAAGGTGAAGAGAAATGGGA	423
Db	361	TCAGGTGATGAGAGCCAAAGCTTAGCTCTCTGAAGGGAAGGTGAAGAGAAATGGGA	420
Qy	424	GAAGGAAAGAGGGGTCTCTCAGGCTCTTCAGAGAGAGGCTCCAAATCCCTCTGAAA	483
Db	421	GAAGGAAAGAGGGGTGACTTCAGGCTCTGAGAGACAGAGCTCCAAATCC---CTTAAA	477
Qy	484	AGAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTTCGGCCCAAGGATGAGAGGATCTC	543
Db	478	AGAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTTCGGCCCAAGGATGAGAGGACCTC	537
Qy	544	ACATCCCGAGTCTAGTACCGTGAAGGGATGATGCGCGGACCGATGGCGGGTTTAAA	603
Db	538	ACATCCCTGAGTCTAGTACGCTGAAGGGATGATGCGCGGACCGATGGCGGGTTTANA	597
Qy	604	TGGCAGCAACCGCAAGCGAGCTGTTAAAGATGAGATCTCATTTCCATCATCTTCCTG	663
Db	598	TGGCAGCAACCGGACCGAGCCAGTCAAGATGAGATCTCATTTTCATCATTTTCTG	657
Qy	664	CTAAGCCAAACAGAGAGCGAGCTGTTGTACAGGCAATCAGAGAAAGTTGAAATA	723
Db	658	TTGAGCCAAACAGAGAGCGAGCTGTTGTACAGGCAATCAGAGAAAGTTGAAATA	717
Qy	724	TGTGAAGA 731	
Db	718	TGTGAAGA 725	

RESULT 13	
LOCUS	BQ892009
DEFINITION	BQ892009 899 bp mRNA linear EST 16-AUG-2002
ACCESSION	AGENCOURT_8754837 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6391577
VERSION	5', mRNA sequence.
KEYWORDS	EST.
SOURCE	BQ892009.1 GI:22284023
ORGANISM	Mus musculus (house mouse)
REFERENCE	Mus musculus
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
JOURNAL	1 (bases 1 to 899)
COMMENT	NIH-MGC http://mgi.nci.nih.gov/ . National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished Contact: Robert Strausberg, Ph.D. Email: cgabs-r@mail.nih.gov Tissue Procurement: Susan L. Sullivan, PhD. cDNA Library Preparation: Resgen, Invitrogen Corp cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Plate: L1AM1380 Row: a Column: 18
 High quality sequence stop: 683.
 Location/Qualifiers
 source
 1. .899
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:6391577"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH MGC 129"
 /notes="Organ: olfactory epithelium; Vector: pCMV-SPORT6.1.cdb; Site 1: EcoRV; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2.2 kb. Constructed by Resgen, Invitrogen Corp. Note: this is a NIH MGC Library."
 BASE COUNT 277 a 190 c 257 g 174 t 1 others
 ORIGIN

Query Match 84.8%; Score 633.4; DB 13; Length 899;
 Best Local Similarity 92.8%; Pred. No. 5.2e-156;
 Matches 677; Conservative 0; Mismatches 51; Indels 3; Gaps 1;
 QY 1 ATGGCGGCTGTGAGTCTGCGGCTCGGCGACTTGGTGTGGGGAACTCGGCCGATATCCT 60
 DB 5 ATGGCGGCTGTGAGTCTGCGGCTCGGCGACTTGGTGTGGGGAACTCGGCCGATATCCT 64
 QY 61 CTTGGCCAGGAAAGATTGTTAATCCAAAGGACTTGAAGAAACCTCGCGGAAGAA 120
 DB 65 CCTGGCCAGGAAAGATTGTTAATCCAAAGGACTTGAAGAAACCTCGCGGAAGAA 124
 QY 121 TCGTCTTTGTGAAATTTTTCGACAGAGATCATGCTGGATCAAGTGGACAGACTG 180
 DB 125 TCGTCTTTGTGAAATTTTTCGACAGAGATCATGCTGGATCAAGTGGACAGACTG 184
 QY 181 AAGCCATATCATGCTCATAAAGAGGAAATGATAAATAATTAACAAGGGTAAACGATTCCAG 240
 DB 185 AAGCCTTACCTGCTCACAAGGAGGAGATGATAAGATTAAACAAGGGTAAACGATTCCAG 244
 QY 241 CAAGCGGTAGATGCTGTCTCAAGAGTTCCTCAGAGAGCCAAAGGGAAAGACAGAGCTCA 300
 DB 245 CAAGCTGTGATGCTGTGTAAGAGTTCCTCAGAGAGCCAAAGGGAAAGACAGACATCA 304
 QY 301 TCCCAATTTCTTGATGACAGAGATCGACGTAATTCAGTCAAGAGAGAGTAGGCCCA 360
 DB 305 TCCCACTTCTGCTGATGACAGAAATCGGCGTAATTCAGTGAAGAGAGTAGGCCCA 364
 QY 361 AACTCAGGTGATGAGAAAGCGCAAACTTAGCTGTCTGAAGGGAGGTGAAGAGAACATG 420
 DB 365 AACTCAGGTGATGAGAAAGCGCAAACTTAGCTGTCTGAAGGGAGGTGAAGAGAACATG 424
 QY 421 GGAGAGGAAAGAGAGGTGTCTTCAGGCTCTTCAGAGAGAGGTTCCTCAATCCCTCTG 480
 DB 425 GGAGAGGAAAGAGAGGTGTCTTCAGGCTCTTCAGAGAGAGGTTCCTCAATCCCTCTG 481
 QY 481 AAAAGAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTTCGCCCCCAAGAGTAGAGAGAT 540
 DB 482 AAAAGAGCCCAAGAGCAAAAGTCCCGGAAGCGGGTTCGCCCCCAAGAGTAGAGAGAT 541
 QY 541 CTCACCATCCCGAGTCTAGTACCGTGAAGGGAGTATGCGCGGACCGATGGCGGGTTT 600
 DB 542 CTCACCATCCCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 601
 QY 601 AAATGCGCCAAACCGCAAGCGAGCTGTGTAAGATGAGATGAGATGAGATGAGATGAG 660
 DB 602 AAATGCGCCAAACCGCAAGCGAGCTGTGTAAGATGAGATGAGATGAGATGAGATGAG 661
 QY 661 CTGCTAAGCCAAACAGAGAGAGCAGTGTGTTACCGGCAATCAGAGAGAGTTGAA 720
 DB 662 CTGTTAGCCCAACAGAGAGAGCAGTGTGTTACCGGCAATCAGAGAGAGTTGAA 721

721 ATATGTGAAGA 731
 722 ATATGTGAAGA 732

RESULT 14
 BQ918182
 LOCUS
 DEFINITION BQ918182 890 bp mRNA linear EST 20-AUG-2002
 5', mRNA sequence.
 AGENCOURT_8765993 NIH_MGC_130 Mus musculus cDNA clone IMAGE:6330861
 BQ918182
 VERSION
 BQ918182.1 GI:22332892
 EST.
 SOURCE
 Mus musculus (house mouse)
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (base 1 to 890)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE
 National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-x@mail.nih.gov
 Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman,
 Ph.D.
 cDNA Library Preparation: ResGen, Invitrogen Corp
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
 Plate: L1AM13785 row: O column: 22
 High quality sequence stop: 648.

FEATURES
 source
 1..890
 Location/Qualifiers
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:6330861"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH MGC 130"
 /note="Organ: oocytes; Vector: pCMV-SPORT6.1.cdb;
 Site 1: EcoRV; Site 2: NotI; Cloned unidirectionally.
 Primer: Oligo dT. Average insert size 1.95 kb.
 Constructed by ResGen, Invitrogen Corp. Note: this is a
 NIH MGC Library."
 BASE COUNT 268 a 201 c 245 g 175 t 1 others
 ORIGIN

Query Match 84.8%; Score 633.2; DB 13; Length 890;
 Best Local Similarity 92.8%; Pred. No. 5.9e-156;
 Matches 675; Conservative 0; Mismatches 49; Indels 3; Gaps 1;

QY 5 CGGCTGTGAGTCTGCGGCTCGGCGACTGTGTGGGGAAATCGCGCGATATCTCCCTT 64
 Db 1 CGGCTGTGAGTCTGCGGCTCGGCGACTGTGTGGGGAAATCGCGCGATATCTCCCTT 60
 QY 65 GGCAGGAAAGATGTTAATCCCAAGGACTTGAAGAACTCCGCGAAAGAAATGCT 124
 Db 61 GGCAGGAAAGATGTTAATCCCAAGGACTTGAAGAACTCCGCGAAAGAAATGCT 120
 QY 125 TCTTTGTGAATTTTGGAAACAGAGATCATGCTCGATCAAAAGTGAACAGCTGAAGC 184
 Db 121 TCTTTGTGAATTTTGGAAACAGAGATCATGCTCGATCAAAAGTGAACAGCTGAAGC 180
 QY 185 CATATCATGCTATAAAGAGAAATGATAAAATTAACAGGGTAAACAGATTCAGCAAG 244
 Db 181 CTTACCATGCTCAAGAGGAGATGATAAAGATTAAAGGTTAAACAGGTTCCAGCAAG 240
 QY 245 CGGTAGATGCTGTGGAAGATTTCTCAGGAGAGCCAAAGGAAAGACCAAGCTATCCC 304
 Db 241 CTTGGAATGCTGTGGAAGATTTCTCAGGAGAGCCAAAGGAAAGACCAAGCTATCCC 300

QY 305 ACAATTTCTTCATGACAAAGATCGACGTAATTCAGTCAGGAGAGAACTAGGCCAAACT 364
 Db 301 ACACTTTCTTCATGACAAAGATCGGCGTAATTCAGTCAGGAGAGAACTAGGCCAAACT 360
 QY 365 CAGGTGATGAGAAGCGCAAACTTAGCTCTCTGAAGGGAGGTGAAGAAGAACTGGAG 424
 Db 361 CAGGTGATGAGAAGCGCAAGCTTAGCTCTCTGAAGGGAGGTGAAGAAGAACTGGAG 420
 QY 425 AAGGAAAGAGAGGGTGTCTTCAGGCTTTCAGAGAGAGGCTCCAAATCCCTCTCAAAA 484
 Db 421 AAGGAAAGAGAGGGTGTCTTCAGGCTTTCAGGCTTTCAGAGAGAGGCTCCAAATGC---CTTAAA 477
 QY 485 GAGCCCAAGAGCAAAAGTCCCGGAGAGCGGGTCCGCCCCCAAGAGATGAGAAGATCTCA 544
 Db 478 GAGCCCAAGAGCAAAAGTCCCGGAGAGCGGGTCCGCCCCCAAGAGATGAGAAGATCTCA 537
 QY 545 CCATCCCGAGTCTAGTACCGTGAAGGGGATGATGCGCGACCGATGGCGCGTCTTAAAT 604
 Db 538 CCATCCCTGAGTCTAGTCTGTAAGGGGATGATGCTGACCGATGGCTGCAATTTAAAT 597
 QY 605 GGCGACCAACCGCAGCGAGCCTGTTTAAAGATGCGAGATCCTCATTTCCATCATTTCTGC 664
 Db 598 GGCGACCAACCGCAGCGAGCCTGTTTAAAGATGCGAGATCCTCATTTCTCATTTCTGT 657
 QY 665 TTAGCCAAACAGAGAGCCAGCTGTCTGTACAGCAATCAGCAAGAGTTGCAAAATAT 724
 Db 658 TTAGCCAAACAGAGAGCCAGCTGTCTGTACAGCAATCAGCAAGAGTTGCAAAATAT 717
 QY 725 GTGAAGA 731
 Db 718 GTGAAGA 724

RESULT 15
 LOCUS
 AK014456 3239 bp mRNA linear HTC 05-DEC-2002
 DEFINITION Mus musculus 13 days embryo whole body cDNA, RIKEN full-length
 enriched library, clone:3930401K13 product:CYTOKINE-LIKE NUCLEAR
 FACTOR N-PAC homolog [Homo sapiens], full insert sequence.

ACCESSION AK014456
 VERSION AK014456.1 GI:12852319
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
 1 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning
 Meth. Enzymol. 303, 19-44 (1999)
 JOURNAL 99279253
 MEDLINE
 PUBMED 10349636

REFERENCE
 2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)
 JOURNAL 20499374
 MEDLINE
 PUBMED 11042159

REFERENCE
 3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 Konno, H., Akiyama, J., Nishi, K., Kitsumai, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Fujisawa, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 JOURNAL 20530913
 MEDLINE
 PUBMED 11076861

QY	601	AAATGGCAGCCACCGAGCGAGCCTGTTAAAGATGCAGATCCTCATTTCCATCATTTTC	660
Db	605	AAATGGCAGCCAAACAGCGACCGAGCCAGTCAAAAGATGCAGATCCTCATTTTCATCATTTT	664
QY	661	CTGCTAAGCCAAACAGAGAGCCAGCTGTCTGTTACCGCAATCACGAAGATTGAAA	720
Db	665	CTGTTGAGCCAAACAGAGAGCCAGCTGTCTGTTACCGCAATCACGAAGATTGAAA	724
QY	721	ATATGTGAAGA	731
Db	725	ATATGTGAAGA	735

Search completed: December 8, 2003, 14:55:59
 Job time : 2083 secs

THIS PAGE BLANK (USPTO)